# The origin of black hole entropy

Shinji Mukohyama
Yukawa Institute for Theoretical Physics, Kyoto University
Kyoto 606-8502, Japan

Doctoral thesis submitted to
Department of Physics, Kyoto University
December 1998

**Abstract**

In this thesis properties and the origin of black hole entropy are investigated from various points of view. First, laws of black hole thermodynamics are reviewed. In particular, the first and generalized second laws are investigated in detail. It is in these laws that the black hole entropy plays key roles. Next, three candidates for the origin of the black hole entropy are analyzed: the D-brane statistical-mechanics, the brick wall model, and the entanglement thermodynamics. Finally, discussions are given on semiclassical consistencies of the brick wall model and the entanglement thermodynamics and on the information loss problem.

# Contents

# Chapter 1

# Introduction

Thermodynamics describes behavior of coarse-grained or averaged quantities of a system with a large number of physical degrees of freedom. The behavior is traced by a small number of parameters. Mathematically, the microscopic description of thermodynamics, statistical mechanics, is grounded by the ergodic hypothesis. On the other hand, in the theory of black holes, the no hair theorem [1] allows us to describe a stationary black hole by a small number of parameters. Since the cosmic censorship conjecture [2] combined with the singularity theorem [3] predicts inevitable occurrence of black holes, it seems that the no hair theorem plays the same role as the ergodic hypothesis plays in thermodynamics.

In fact, it is well known that black holes have many properties analogous to those of thermodynamics. Those are as a whole called *black hole thermodynamics*. In particular, four laws of black holes combined with the generalized second law make up a main framework of the black hole thermodynamics. In these laws, *black hole entropy* defined as follows plays an important role.

$$S_{BH} = \frac{k_B c^3}{4\hbar G} A, \tag{1.1}$$

where $A$ is area of black hole horizon. Moreover, it was suggested that the black hole entropy, or the horizon area, is an adiabatic invariant [4] and that it can be used as a potential function in catastrophe theory to judge stability of black hole solutions [5]. The formula (1.1) for black hole entropy is often called *Bekenstein-Hawking formula* since the concept of black hole entropy was first introduced by Bekenstein [6] as a quantity proportional to the horizon area and the proportionality coefficient was fixed by Hawking's discovery of thermal radiation from a black hole [7] (see arguments below). He showed that a black hole radiates thermal radiation with temperature given by

$$k_B T_{BH} = \frac{\hbar \kappa}{2\pi c}, \tag{1.2}$$

where $\kappa$ is the surface gravity of a background black hole. This thermal radiation and its temperature are called *Hawking radiation* and *Hawking temperature*, respectively.

Let us recall basic properties of the black-hole thermodynamics by taking a simple example. We consider a one-parameter family of Schwarzschild black holes parameterized by the mass $M_{BH}$. We assume that a relation analogous to the first law of thermodynamics holds for a black-hole system. In the present example, there is only one parameter $M_{BH}$ characterizing a black hole. Therefore, this relation should be of the simplest form

$$\delta E_{BH} = T_{BH} \delta S_{BH}, \tag{1.3}$$

where $E_{BH}$, $S_{BH}$ and $T_{BH}$ are quantities that are identified with the energy, the entropy and the temperature of a black hole, respectively. The relation Eq.(1.3) is called *the first law* of the black-hole thermodynamics [8]. Thus, if two of the quantities $E_{BH}$, $S_{BH}$ and $T_{BH}$ are given, *Eq.*(1.3) determines the remaining quantity.

In the present example, $M_{BH}$ is the only parameter characterizing the family of black holes. Therefore the simplest combination which yields the dimension of energy is

$$E_{BH} \equiv M_{BH}c^2. \tag{1.4}$$

This is the energy of the black hole.

There is also a natural choice for $T_{BH}$ [7]. Hawking showed that a black hole with surface gravity $\kappa$ emits thermal radiation of a quantum matter field (which plays the role of a thermometer) at temperature given by (1.2). Moreover, as shown in section 2.3, if a matter field in a thermal-equilibrium state at some temperature is scattered by a black hole, then it always becomes closer to the thermal-equilibrium state at the Hawking temperature (1.2) [9, 10]. Thus it is natural to define the temperature of a Schwarzschild black hole with mass $M_{BH}$ by

$$k_B T_{BH} = \frac{\hbar c^3}{8\pi G M_{BH}}, \tag{1.5}$$

since $\kappa = c^4/4GM_{BH}$ [11].

From Eqs.(1.3)-(1.5), we get an expression for $S_{BH}$ as

$$S_{BH} = \frac{k_B c^3}{4\hbar G} A + C, \tag{1.6}$$

where $A \equiv 16\pi G^2 M_{BH}^2/c^4$ is the area of the event horizon and $C$ is some constant. Since a value of $C$ is not essential in our discussions, we shall set hereafter

$$C = 0. \tag{1.7}$$

Note that Eq. (1.6) with (1.7) is a special case of the Bekenstein-Hawking formula (1.1).

It is well-known that classically the area of the event horizon does not decrease in time (the area law [12] or *the second law* of black hole) just as the ordinary thermodynamical entropy. The Bekenstein-Hawking formula (1.1) looks reasonable in this sense. Indeed this observation was the original motivation for the introduction of the black-hole entropy [6]. Moreover, when quantum effects are taken into account, it is believed that a sum of the black hole entropy and matter entropy does not decrease (*the generalized second law*).

*The zeroth law* of black hole thermodynamics states that surface gravity of a Killing horizon is constant throughout the horizon. This supports our choice of the black hole temperature. (For a proof of the zeroth law, see Refs. [8, 11, 13].)

At this stage we would like to point out that for a black hole the third law does not hold in the sense of Planck: $S_{BH} \to \infty$ as $T_{BH} \to 0$ for the family of Schwarzschild black holes, irrespective of the choice of the value for $C$ as is seen from Eqs.(1.5) and (1.6). Rather, *the third law* does hold in the sense of Nernst: it is impossible by any process, no matter how idealized, to reduce $\kappa$ to zero in a finite sequence of operations [8]. (See Ref. [14] for a precise expression and a proof.)

Thermodynamics has a well-established microscopic description: the quantum statistical mechanics. In the thermodynamical description, information on each microscopic degree of freedom is lost, and only macroscopic variables are concerned. However, the number of all microscopic degrees of freedom is implemented in a macroscopic variable: entropy $S$ is related to the number of all consistent microscopic states $N$ as

$$S = k_B \ln N. \tag{1.8}$$

In analogy, it is expected that there might be a microscopic description of the black hole thermodynamics, too. In particular, it is widely believed that the black hole entropy might be related to a number of microscopic states. Since the microscopic description seems to require a quantum theory of gravity, detailed investigations of the black hole entropy should contribute a lot toward construction of the theory of quantum gravity. This is one among the several reasons why the origin of the black hole entropy needs to be understood at the fundamental level.

Another strong motivation to investigate the black hole entropy is the so-called information loss problem. Hawking argued that, if a black hole is formed by gravitational collapse, then evolution of quantum fields becomes non-unitary because of evaporation of the black hole due to the Hawking radiation [15]. This means that some information is lost in the process of the black hole evaporation. Moreover, this suggests that the conventional field-theoretical approach may be useless for the purpose of construction of the theory of quantum gravity since the field theory is based on the unitarity. Hence, the evaporation of a black hole makes people, who wish to construct a unitary theory of quantum gravity, be in difficulties: if the evaporation of a black hole would actually occur and lead to the information loss, then they would be obliged to give up the unitarity. Thus, we have to clarify whether information is lost or not due to the Hawking radiation in order to take a step forward. This problem is called the *information loss problem*. On the other hand, since entropy is strongly connected with information in the theory of information, it is natural to expect that the black hole entropy might be related to some information. Therefore, investigations of the origin of the black hole entropy seem to provide important insight toward the information loss problem.

In these senses, the origin of the black hole entropy is one of the most important issues at the present stage of black hole physics.

Recently a microscopic derivation of the black hole entropy was given in superstring theory [16, 17] by using the so-called D-brane technology [18]. In this approach, as will be shown in section 3.1, the black hole entropy is identified with the logarithm of the number of states of massless strings attached to D-branes, with D-brane configuration and total momentum of the strings along a compactified direction fixed to be consistent with the corresponding black hole [19, 20]. The analysis along this line was extended to the so-called M-theory [21]. In particular, by using a conjectured correspondence (the Matrix theory) between the M-theory in the infinite momentum frame and a 10-dimensional $U(N)$ supersymmetric Yang-Mills theory dimensionally reduced to $(0+1)$-dimension with $N \to \infty$ [22], the black hole entropy was calculated by means of the Yang-Mills theory. The result gives the correct Bekenstein-Hawking entropy for BPS black holes and their low lying excitations [23]. Moreover, in Ref. [24] the black hole entropy of a Schwarzschild black hole was derived in the Matrix theory up to a constant of order unity. On the other hand, in loop quantum gravity [25], black hole entropy was identified with the logarithm of the number of different spin-network states for a fixed eigenvalue of the area operator [26]. The result coincides with the Bekenstein-Hawking entropy up to a constant of order unity.

The derivations in these candidate theories of quantum gravity depend strongly on details of the theories. In this sense, the success of the derivations can be considered as non-trivial consistency checks of the theories. However, it is believed that proportionality of the black hole entropy to horizon area is more universal and does not depend on details of the theory. Hence, one should be able to give a statistical or thermodynamical derivation of the black hole entropy, which does not depend on details of theory, while we are proceeding with theory-dependent derivations of it by using the well-established candidate theories of quantum gravity.

There were many attempts to explain the origin of the black hole entropy besides the above theory-dependent approaches. (See Ref. [27] for an up-to-date re-

view.) For example, in Euclidean gravity the black hole entropy is associated with the topology of an instanton which corresponds to a black hole [28, 29, 30, 31] [1]; Wald [33] defined the black hole entropy as a Noether charge associated with a bifurcating Killing horizon [2] (See section 2.1.); 'tHooft [35] identified the black hole entropy with the statistical entropy of a thermal gas of quantum particles with a mirror-like boundary just outside the horizon (This model is called the brick wall model and is analyzed in detail in section 3.2); Pretorius et al. [36] identified the black hole entropy with the thermodynamical entropy of a shell in thermal equilibrium with acceleration radiation due to the shell's gravity in the limit that the shell forms a black hole.

There remains another strong candidate for the statistical origin of the black hole entropy, called entanglement entropy [37, 38, 39]. It is a statistical entropy measuring the information loss due to a spatial division of a system [37]. The entanglement entropy is based only on the spatial division, and can be defined independently of the theory, although explicit calculations in the literature are dependent on the model employed. Moreover, as will be explained in section 3.3, it is expected independently of the details of the theory that the entanglement entropy is proportional to the area of the boundary of the spatial division. In this sense, the entanglement entropy is considered to be a strong candidate for the statistical origin of the black hole entropy.

In chapter 2, laws of black hole thermodynamics are reviewed. In particular the first laws of black hole statics and dynamics, and the generalized second law are studied in detail. The first law derived in the above simple arguments on Schwarzschild black holes relates changes of physical quantities of stationary black holes corresponding to a variation in a space of stationary black hole solutions. In this sense we can call it *the first law of black hole statics*. Historically, the first law of black hole statics is derived in Ref. [8] in general relativity, and is extended by Wald to a general covariant theory of gravity [33]. In section 2.1 we re-analyze the first law of Wald in detail following Ref. [40]. In section 2.2 we consider a generalization of the first law to a purely dynamical situation [41, 42]. We call the dynamical version *the first law of black hole dynamics*. In section 2.3 a proof of the generalized second law is given for a quasi-stationary black hole [10].

In chapter 3, three candidates for the origin of the black hole entropy are analyzed in detail. In section 3.1 a microscopic derivation of black hole entropy by the D-brane technology is shown. We consider a 5-dimensional black hole solution in the low energy effective theory of Type IIB superstring. This black hole is, in fact, a black brane in 10-dimensional sense and can be interpreted as a configuration of D-branes wrapped on $T^5 = T^4 \times S^1$. We calculate statistical-mechanical entropy and temperature of open strings on the D-branes and compare them with the Bekenstein-Hawking entropy and the Hawking temperature of the original 5-dimensional black hole [43]. In section 3.2 we re-examine the brick wall model in detail and solve problems concerning this model [44]. In section 3.3 we construct a thermodynamics (entanglement thermodynamics [45, 46]) which includes the entanglement entropy as the entropy variable, for a massless scalar field in Minkowski, Schwarzschild and Reissner-Nordström spacetimes to understand the statistical origin of black-hole thermodynamics. In section 3.4 a new interpretation of entanglement entropy is proposed [47].

Chapter 4 is devoted to a summary of this thesis, discussions on semiclassical consistencies and the information loss problem, and speculations.

[1] The 1-loop correction to the black hole entropy was also calculated and compared with the brick wall model and the conical singularity method [32].

[2] Relations to the approach by Euclidean gravity was investigated in Ref. [34].

# Chapter 2

# Laws of black hole thermodynamics

## 2.1 The first law of black hole statics

In Ref. [33], the first law of black hole mechanics was derived not only in general relativity but also in a general covariant theory of gravity for stationary variations around a stationary black hole. It is formulated as a relation among variations of those quantities such as energy, angular momentum and entropy, each of which is defined in terms of a Noether charge. The first law was extended to non-stationary variations around a stationary black hole in Ref. [48].

The Noether charge form of the first law has many advantages over the original first law of Ref. [8]. For example, it gives a general method to calculate stationary black hole entropy in general covariant theories of gravity [48]; it connects various Euclidean methods for computing black hole entropy [34]; it suggests a possibility of defining entropy of non-stationary black holes [48, 49]; etc.

However, in their derivation there are several issues to be discussed in more detail.

(a) In Ref. [33], unperturbed and perturbed stationary black holes are identified so that horizon generator Killing fields with unit surface gravity coincide in a neighborhood of the horizons and that stationary Killing fields and axial Killing fields coincide in a neighborhood of infinity. This corresponds to taking a certain gauge condition in linear perturbation theory. For a complete understanding of the first law, we have to clarify whether such a gauge condition can be imposed or not. If it can, then we wish to know whether such a gauge condition is necessary. Note that, on the contrary, the original derivation in general relativity by Bardeen, Carter and Hawking [8] is based on a gauge condition such that the stationary Killing fields and the axial Killing fields coincide everywhere on a spacelike hypersurface whose boundary is a union of a horizon cross section and spatial infinity.

(b) In Ref. [48], the first law is extended to non-stationary perturbations around a stationary black hole. In the derivation, change of black hole entropy is calculated on a $(n-2)$-surface, which is a bifurcation surface for an unperturbed black hole, but which is not a cross section of an event (nor apparent) horizon for a perturbed non-stationary black hole in general. Does this mean that black hole entropy would be assigned to a surface which is not a horizon cross section for a non-stationary black hole? It seems more natural to assign

black hole entropy to a horizon cross section also for a non-stationary black hole.

In this section these two issues are discussed and it is concluded that there are no difficulties in the derivation of the Noether charge form of the first law for both stationary and non-stationary perturbations about a stationary black hole. In its course, we give an alternative derivation of the first law based on a variation in which a horizon generator Killing field with unit surface gravity is fixed.

In subsection 2.1.1 gauge conditions are analyzed. In subsection 2.1.2 the first law of black holes is derived for stationary variations around a stationary black hole. In subsection 2.1.3 the derivation is extended to non-stationary variations around a stationary black hole.

### 2.1.1 Gauge conditions

Consider a stationary black hole in $n$-dimensions, which has a bifurcating Killing horizon. Let $\xi^a$ be a generator Killing field of the Killing horizon, which is normalized as $\xi^a = t^a + \Omega_H^{(\mu)} \varphi^a_{(\mu)}$, and $\Sigma$ be the bifurcation surface. Here, $t^a$ is the stationary Killing field with unit norm at infinity, $\{\varphi^a_{(\mu)}\}$ $(\mu = 1, 2, \cdots)$ is a family of axial Killing fields, and $\{\Omega_H^{(\mu)}\}$ is a family of constants (angular velocities). Let $\kappa$ be the surface gravity corresponding to $\xi^a$:

$$\xi^b \nabla_b \xi^a = \kappa \xi^a \tag{2.1}$$

on the horizon.

Now let us show that it is not possible in general to impose a gauge condition such that $\delta\xi^a = 0$ in a neighborhood of the bifurcation surface. For this purpose we shall temporarily assume that $\delta\xi^a = 0$ and show a contradiction.

On $\Sigma$, the covariant derivative of $\xi^a$ is given by

$$\nabla_b \xi^a = \kappa \epsilon_b{}^a, \tag{2.2}$$

where $\epsilon_{ab}$ is binormal to $\Sigma$. However, the variation of the l.h.s. is zero:

$$\delta(\nabla_b \xi^a) = \delta\Gamma^a_{bc} \xi^c = 0 \tag{2.3}$$

since $\xi^a = 0$, where $\delta\Gamma^a_{bc}$ is given by

$$\delta\Gamma^a_{bc} = \frac{1}{2} g^{ad} (\nabla_c \delta g_{db} + \nabla_b \delta g_{dc} - \nabla_d \delta g_{bc}). \tag{2.4}$$

Hence,

$$\delta\epsilon_b{}^a = -\frac{\delta\kappa}{\kappa} \epsilon_b{}^a. \tag{2.5}$$

Substituting this into the identity $\delta(\epsilon_b{}^a \epsilon^b{}_a) = 0$, we obtain

$$0 = \delta(\epsilon_b{}^a \epsilon^b{}_a) = -\frac{4\delta\kappa}{\kappa}. \tag{2.6}$$

Thus, the assumption $\delta\xi^a = 0$ leads to $\delta\kappa = 0$, which implies, for example, that $\delta M = 0$ for the vacuum general relativity in a static case, where $M$ is mass of Schwarzschild black holes. This peculiar behavior can be understood as appearance of a coordinate singularity at the bifurcation surface of a coordinate fixed by the gauge condition $\delta\xi^a = 0$ since in the above argument finiteness of $\delta\Gamma^a_{bc}$ has been assumed implicitly. Therefore, it is impossible to impose the condition $\delta\xi^a = 0$ in a neighborhood of the bifurcation surface whenever $\delta\kappa \neq 0$.

As mentioned above, the original derivation of the first law in Ref. [8] adopt the gauge condition $\delta t^a = \delta\varphi^a = 0$. This leads to $\delta\xi^a = 0$ when $\delta\Omega = 0$ (for example, when we consider static black holes). Of course, in Ref. [8], a general horizon cross section (not necessary a bifurcation surface) is considered as a surface on which black hole entropy is calculated. Hence, the above argument arises no difficulties unless the cross section is taken to be the bifurcation surface. The derivation in Ref. [8] suffers from the above argument if and only if black hole entropy is estimated on the bifurcation surface.

On the other hand, arguments like the above do not lead to any contradiction if we adopt a gauge condition such that $\tilde{\xi}^a$ is fixed in a neighborhood of the bifurcation surface under variations, where $\tilde{\xi}^a = \xi^a/\kappa$ is a horizon generator Killing field with unit surface gravity. Moreover, it is concluded that, if we intend to fix a horizon generator Killing field, then it must have the same value of surface gravity for unperturbed and perturbed black holes. Hence, the gauge condition $\delta\tilde{\xi}^a = 0$ in a neighborhood of the bifurcation surface adopted in Ref. [33, 48] is very natural one.

In fact, it is always possible to identify unperturbed and perturbed stationary black holes so that the Killing horizons and the generator Killing fields with unit surface gravity coincide. As stated in Ref. [33], such an identification can be done at least in a neighborhood of the horizon by using the general formula for Kruskal-type coordinates $(U, V)$ given in Ref. [13]. (The identified Killing horizon is given by $U = 0$ and $V = 0$. The identified Killing field with unit surface gravity is given by $\tilde{\xi}^a = U(\partial/\partial U)^a - V(\partial/\partial V)^a$.)

The purpose of the next subsection is to discuss the remaining gauge condition $\delta t^a = \delta\varphi^a = 0$ at infinity. It is evident that this gauge condition at infinity can be imposed by identifying the perturbed and unperturbed spacetimes suitably. So, our question now is whether this gauge condition is necessary or not. For this purpose we temporarily adopt a gauge condition such that $\tilde{\xi}^a$ is fixed everywhere on a hypersurface connecting the bifurcation surface and spatial infinity. In deriving the first law in this gauge condition, the gauge condition $\delta t^a = \delta\varphi^a = 0$ at infinity is found to be necessary for a proper interpretation of the first law. On the other hand, as shown in subsection 2.1.3, it is not necessary to fix $\tilde{\xi}^a$ in a neighborhood of the bifurcation surface, strictly speaking. Hence, it can be concluded that the minimal set of gauge conditions necessary for the derivation of the first law is that $t^a$ and $\varphi^a$ are fixed at spatial infinity.

### 2.1.2 The first law for stationary black holes

Before deriving the first law, we review basic ingredients of the formalism.

We consider a classical theory of gravity in $n$-dimensions with arbitrary matter fields, which is described by a diffeomorphism invariant Lagrangian $n$-form $\mathbf{L}(\phi)$, where $\phi$ denotes dynamical fields [48].

The Noether current $(n-1)$-form $\mathbf{j}[V]$ for a vector field $V^a$ is defined by

$$\mathbf{j}[V] \equiv \mathbf{\Theta}(\phi, \mathcal{L}_V\phi) - V \cdot \mathbf{L}(\phi), \tag{2.7}$$

where the $(n-1)$-form $\mathbf{\Theta}(\phi, \delta\phi)$ is defined by

$$\delta\mathbf{L}(\phi) = \mathbf{E}(\phi)\delta\phi + d\mathbf{\Theta}(\phi, \delta\phi). \tag{2.8}$$

It is easily shown that the Noether current is closed as

$$d\mathbf{j}[V] = -\mathbf{E}(\phi)\mathcal{L}_V\phi = 0, \tag{2.9}$$

where we have used the equations of motion $\mathbf{E}(\phi) = 0$. Hence, by using the machinery developed in Ref. [50], we obtain the Noether charge $(n-2)$-form $\mathbf{Q}[V]$ such that

$$\mathbf{j}[V] = d\mathbf{Q}[V]. \tag{2.10}$$

Hereafter, we assume that in an asymptotically flat spacetime there exists an $(n-1)$-form $\mathbf{B}$ such that

$$\int_\infty V \cdot \delta\mathbf{B}(\phi) = \int_\infty V \cdot \mathbf{\Theta}(\phi, \delta\phi), \tag{2.11}$$

where the integral is taken over an $(n-2)$-dimensional sphere at infinity. By using $\mathbf{B}$, we can write a Hamiltonian $H[V]$ corresponding to evolution by $V^a$ as follows [33].

$$H[V] \equiv \int_\infty (\mathbf{Q}[V] - V \cdot \mathbf{B}). \tag{2.12}$$

The symplectic current density $\omega(\phi, \delta_1\phi, \delta_2\phi)$ is defined by

$$\omega(\phi, \delta_1\phi, \delta_2\phi) \equiv \delta_1[\mathbf{\Theta}(\phi, \delta_2\phi)] - \delta_2[\mathbf{\Theta}(\phi, \delta_1\phi)] \tag{2.13}$$

and is linear both in $\delta_1\phi$ and its derivatives, and $\delta_2\phi$ and its derivatives [51].

Now we define a space of solutions in which we take a variation to derive the first law.

Let $\tilde{\xi}^a$ be a fixed vector field, which vanishes on a $(n-2)$-surface $\Sigma$. (Note that $\tilde{\xi}^a$ and $\Sigma$ can be defined without referring to any dynamical fields, eg. the metric $g_{ab}$.) In the following arguments, we consider a space of stationary, asymptotically flat solutions of the equations of motion $\mathbf{E}(\phi) = 0$, each of which satisfies the following three conditions. (a) There exists a bifurcating Killing horizon with the bifurcation surface $\Sigma$. (b) $\tilde{\xi}^a$ is a generator Killing field of the Killing horizon. (c) Surface gravity corresponding to $\tilde{\xi}^a$ is 1:

$$\tilde{\xi}^b \nabla_b \tilde{\xi}^a = \tilde{\xi}^a \tag{2.14}$$

on the Killing horizon.

For each element in this space, there exist constants $\kappa$ and $\Omega_H^{(\mu)}$ $(\mu = 1, 2, \cdots)$ such that

$$\kappa\tilde{\xi}^a = t^a + \Omega_H^{(\mu)} \varphi^a_{(\mu)}, \tag{2.15}$$

where $t^a$ is the stationary Killing field with unit norm at infinity, $\{\varphi^a_{(\mu)}\}$ $(\mu = 1, 2, \cdots)$ is a family of axial Killing fields. Hence, $\kappa$ is surface gravity and $\Omega_H^{(\mu)}$ are angular velocities of the horizon.

Note that, by definition, the vector field $\tilde{\xi}^a$ is fixed under a variation of dynamical fields. We express this explicitly by denoting the variation by $\tilde{\delta}$:

$$\tilde{\delta}\tilde{\xi}^a = 0. \tag{2.16}$$

We now derive the first law of black hole mechanics.

First, by taking a variation of the definition (2.7) for $\mathbf{j}[\tilde{\xi}]$ and using (2.16) and (2.8), we obtain

$$\begin{aligned} \tilde{\delta}\mathbf{j}[\tilde{\xi}] &= \tilde{\delta}\left(\mathbf{\Theta}(\phi, \mathcal{L}_{\tilde{\xi}}\phi)\right) - \tilde{\xi} \cdot \left(\mathbf{E}(\phi)\tilde{\delta}\phi + d\mathbf{\Theta}(\phi, \tilde{\delta}\phi)\right) \\ &= \omega(\phi, \tilde{\delta}\phi, \mathcal{L}_{\tilde{\xi}}\phi) + d\left(\tilde{\xi} \cdot \mathbf{\Theta}(\phi, \tilde{\delta}\phi)\right). \end{aligned} \tag{2.17}$$

Here we have used the equations of motion $\mathbf{E}(\phi) = 0$ and the following identity for an arbitrary vector $V^a$ and an arbitrary differential form $\mathbf{\Lambda}$ to obtain the last line.

$$\mathcal{L}_V \mathbf{\Lambda} = V \cdot d\mathbf{\Lambda} + d(V \cdot \mathbf{\Lambda}). \tag{2.18}$$

Since $\omega(\phi, \tilde{\delta}\phi, \mathcal{L}_{\tilde{\xi}}\phi)$ is linear in $\mathcal{L}_{\tilde{\xi}}\phi$ and its derivatives, we obtain

$$d(\tilde{\delta}\mathbf{Q}[\tilde{\xi}]) = d\left(\tilde{\xi} \cdot \mathbf{\Theta}(\phi, \tilde{\delta}\phi)\right) \tag{2.19}$$

by using $\mathcal{L}_{\tilde{\xi}}\phi = 0$ and Eq. (2.10).

Next we ingrate Eq. (2.19) over an asymptotically flat spacelike hypersurface $\mathcal{C}$, which is parallel to $\varphi^a_{(\mu)}$ at infinity and the interior boundary of which is $\Sigma$. Since $\tilde{\xi}^a = 0$ on $\Sigma$, we obtain

$$\tilde{\delta}\int_\Sigma \mathbf{Q}[\tilde{\xi}] = \tilde{\delta}H[\tilde{\xi}]. \tag{2.20}$$

Finally we rewrite the r.h.s. and the l.h.s. of (2.20) in a form useful to be estimated at infinity and the horizon, respectively.

A relation among variations of $\kappa$, $\Omega^{(\mu)}_H$, $t^a$ and $\varphi^a_{(\mu)}$ is obtained by substituting (2.15) in (2.16).

$$t^a\tilde{\delta}\left(\frac{1}{\kappa}\right) + \varphi^a_{(\mu)}\tilde{\delta}\left(\frac{\Omega^{(\mu)}_H}{\kappa}\right) = -\frac{1}{\kappa}\tilde{\delta}t^a - \frac{\Omega^{(\mu)}_H}{\kappa}\tilde{\delta}\varphi^a_{(\mu)}. \tag{2.21}$$

By using this relation and the fact that $H[V]$ is linear in the vector field $V$, we can rewrite the r.h.s. of (2.20) as follows.

$$\begin{aligned}\tilde{\delta}H[\tilde{\xi}] &= \frac{1}{\kappa}(\tilde{\delta}H[t] - H[\tilde{\delta}t]) + \frac{\Omega^{(\mu)}_H}{\kappa}(\tilde{\delta}H[\varphi_{(\mu)}] - H[\tilde{\delta}\varphi_{(\mu)}]) \\ &= \frac{1}{\kappa}\delta_\infty H[t] + \frac{\Omega^{(\mu)}_H}{\kappa}\delta_\infty H[\varphi_{(\mu)}],\end{aligned} \tag{2.22}$$

where the variation $\delta_\infty$ is defined for linear functionals $F[t]$ and $G_{(\mu)}[\varphi_{(\mu)}]$ so that

$$\begin{aligned}\delta_\infty F[t] &= \tilde{\delta}F[t] - F[\tilde{\delta}t], \\ \delta_\infty G_{(\mu)}[\varphi_{(\mu)}] &= \tilde{\delta}G_{(\mu)}[\varphi_{(\mu)}] - G_{(\mu)}[\tilde{\delta}\varphi_{(\mu)}].\end{aligned} \tag{2.23}$$

This newly introduced variation corresponds to a variation at infinity such that $t^a$ and $\varphi^a$ are fixed:

$$\delta_\infty t^a = \delta_\infty \varphi^a_{(\mu)} = 0. \tag{2.24}$$

In Ref. [48] a useful expression of the Noether charge was given as follows.

$$\mathbf{Q}[V] = \mathbf{W}_c(\phi)V^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}V_{d]} + \mathbf{Y}(\phi, \mathcal{L}_V\phi) + d\mathbf{Z}(\phi, V), \tag{2.25}$$

where $\mathbf{W}_c$, $\mathbf{X}^{cd}$, $\mathbf{Y}$ and $\mathbf{Z}$ are locally constructed covariant quantities. In particular, $\mathbf{Y}(\phi, \mathcal{L}_V\phi)$ is linear in $\mathcal{L}_V\phi$ and its derivatives, and $\mathbf{X}^{cd}$ is given by

$$\left(\mathbf{X}^{cd}(\phi)\right)_{c_3\cdots c_n} = -E^{abcd}_R\epsilon_{abc_3\cdots c_n}. \tag{2.26}$$

Here $E^{abcd}_R$ is the would-be equations of motion form [48] for the Riemann tensor $R_{abcd}$ and $\epsilon_{abc_3\cdots c_n}$ is the volume $n$-form.

By using the form of $\mathbf{Q}$ we can rewrite the integral in the l.h.s. of (2.20) as

$$\int_\Sigma \mathbf{Q}[\tilde{\xi}] = \int_\Sigma \mathbf{X}^{cd}(\phi)\nabla_{[c}\tilde{\xi}_{d]}, \tag{2.27}$$

where we have used the Killing equation $\mathcal{L}_{\tilde{\xi}}\phi = 0$ and the fact that $\tilde{\xi}^a = 0$ on $\Sigma$.

Using the relation

$$\nabla_c\tilde{\xi}_d = \epsilon_{cd} \tag{2.28}$$

on $\Sigma$, for any stationary solutions we can eliminate explicit dependence of Eq. (2.27) on $\tilde{\xi}$, where $\epsilon_{cd}$ is the binormal to $\Sigma$. Hence, at least within the space of stationary solutions, we can take the variation $\tilde{\delta}$ of the integral without any difficulties.

Thus, we obtain the first law for stationary black holes by rewriting Eq. (2.20) as

$$\frac{\kappa}{2\pi}\tilde{\delta}S = \delta_\infty \mathcal{E} - \Omega_H^{(\mu)}\delta_\infty \mathcal{J}_{(\mu)}, \tag{2.29}$$

where entropy $S$ is defined by

$$S \equiv 2\pi \int_\Sigma \mathbf{X}^{cd}(\phi)\epsilon_{cd}, \tag{2.30}$$

and energy $\mathcal{E}$ and angular momenta $\mathcal{J}_{(\mu)}$ are defined by

$$\begin{aligned} \mathcal{E} &\equiv H[t] = \int_\infty (\mathbf{Q}[t] - t\cdot\mathbf{B}), \\ \mathcal{J}_{(\mu)} &\equiv -H[\varphi_{(\mu)}] = -\int_\infty \mathbf{Q}[\varphi_{(\mu)}]. \end{aligned} \tag{2.31}$$

Note that, in the r.h.s. of Eq. (2.29), variations of $\mathcal{E}$ and $\mathcal{J}_{(\mu)}$ are taken with $t^a$ and $\varphi^a_{(\mu)}$ fixed. This condition is explicitly implemented by the definition (2.23) of $\delta_\infty$ and is necessary for a proper interpretation of the first law.

We conclude this subsection by giving another expression of the entropy.

Since $\tilde{\xi}^a$ is a generator Killing field of the Killing horizon, we have $\mathcal{L}_{\tilde{\xi}}\phi = 0$ and the pull-back of $\tilde{\xi}\cdot\mathbf{L}(\phi)$ to the horizon vanishes. Hence, the definition (2.7) says that the pull-back of $\mathbf{j}[\tilde{\xi}]$ to the horizon is zero [49]. Thus, the integral of $\mathbf{Q}[\tilde{\xi}]$ is independent of the choice of the horizon cross section.

Moreover, it can be shown that the integral in (2.30) does not change even if we replace the integration surface $\Sigma$ by an *arbitrary* horizon cross section $\Sigma'$ [49]. Therefore we obtain

$$S = 2\pi \int_{\Sigma'} \mathbf{X}^{cd}(\phi)\epsilon'_{cd}, \tag{2.32}$$

where $\epsilon'_{cd}$ denotes the binormal to $\Sigma'$.

### 2.1.3 Non-stationary perturbation

In this subsection, we shall derive the first law for a non-stationary perturbation about a stationary black hole with a bifurcating Killing horizon. Unfortunately, for non-stationary perturbations, $\delta\kappa$ and $\delta\Omega_H^{(\mu)}$ do not have meaning of perturbations of surface gravity and angular velocity of the Killing horizon, even if they are defined. However, since the first law (2.29) does not refer to $\delta\kappa$ and $\delta\Omega_H^{(\mu)}$ but only to the unperturbed values of $\kappa$ and $\Omega_H^{(\mu)}$, we expect that the first law holds also for non-stationary perturbations. In the following, we shall show that it does hold.

First, we specify a space of solutions in which we take a variation.

Let $\tilde{\xi}^a_0$ be a fixed vector field, which vanishes on an fixed $(n-2)$-surface $\Sigma$. In this subsection, we consider a space of asymptotically flat solutions of the field equation $\mathbf{E}(\phi) = 0$, for each of which $\tilde{\xi}^a_0$ is an asymptotic Killing field.

For each solution in this space, there exist constants $\kappa$ and $\Omega_H^{(\mu)}$ $(\mu = 1, 2, \cdots)$ such that at spatial infinity

$$\kappa\tilde{\xi}^a_0 = t^a + \Omega_H^{(\mu)}\varphi^a_{(\mu)}, \tag{2.33}$$

where $t^a$ is a timelike asymptotic Killing field with unit norm at infinity, $\{\varphi^a_{(\mu)}\}$ $(\mu = 1, 2, \cdots)$ is a family of axial asymptotic Killing fields orthogonal to $t^a$ at infinity and $\{\Omega_H^{(\mu)}\}$ is a family of constants. Note that the constants $\kappa$ and $\Omega_H^{(\mu)}$ do not have meaning of surface gravity and angular velocities unless we consider

a stationary solution. Moreover, in general, $\tilde{\xi}_0^a$ and $\Sigma$ have no meaning but an asymptotic Killing field and a fixed $(n-2)$-surface, respectively.

Note that, by definition, the vector field $\tilde{\xi}_0^a$ is fixed under the variation. We denote the variation by $\tilde{\delta}$:

$$\tilde{\delta}\tilde{\xi}_0^a = 0. \tag{2.34}$$

On the contrary, $t^a$, $\varphi^a_{(\mu)}$, $\kappa$ and $\Omega_H^{(\mu)}$ are not fixed under the variation since definitions of them refer to dynamical fields, which are varied. Their variations are related by (2.21).

Suppose that an element $\phi_0$ of the space of solutions satisfies the following three conditions. (a') $\phi_0$ is a stationary solution with a bifurcating Killing horizon with the bifurcation surface $\Sigma$. (b') $\tilde{\xi}_0^a$ is a generator Killing field of the Killing horizon of $\phi_0$. (c') Surface gravity of $\phi_0$ corresponding to $\tilde{\xi}_0^a$ is 1:

$$\tilde{\xi}_0^b \nabla_b \tilde{\xi}_0^a = \tilde{\xi}_0^a \tag{2.35}$$

on the Killing horizon.

Now we derive the first law for the *non-stationary* perturbation $\tilde{\delta}\phi$ about the stationary solution $\phi_0$.

First, we mention that the validity of Eq. (2.20) in the previous section depends on the following three facts. (i) The equations of motion $\mathbf{E}(\phi) = 0$ hold for both unperturbed and perturbed fields. (Unless they hold also for perturbed fields, $\tilde{\delta}\mathbf{j}$ can not be rewritten as $d(\tilde{\delta}\mathbf{Q})$.) (ii) $\tilde{\xi}^a$ (corresponding to $\tilde{\xi}_0^a$) is a Killing field of the unperturbed solution. (iii) $\tilde{\xi}^a = 0$ (corresponding to $\tilde{\xi}_0^a = 0$) on $\Sigma$ for unperturbed solution.

These three are satisfied for the unperturbed solution $\phi_0$ and the non-stationary variation $\tilde{\delta}\phi$ around $\phi_0$, too. Thus, Eq. (2.20) is valid, provided that $\tilde{\xi}^a$ is replaced by $\tilde{\xi}_0^a$.

Since $\tilde{\delta}t^a$, $\tilde{\delta}\varphi^a_{(\mu)}$, $\tilde{\delta}\kappa$ and $\tilde{\delta}\Omega_H^{(\mu)}$ are related by Eq. (2.21), we can transform the r.h.s. of (2.20) to obtain

$$\kappa\tilde{\delta}\int_\Sigma \mathbf{Q}[\tilde{\xi}_0] = \delta_\infty \mathcal{E} - \Omega_H^{(\mu)}\delta_\infty \mathcal{J}_{(\mu)}, \tag{2.36}$$

where, as in the previous section, energy $\mathcal{E}$ and angular momenta $\mathcal{J}_{(\mu)}$ are defined by (2.31), and the variation $\delta_\infty$ is defined at infinity so that $t^a$ and $\varphi^a_{(\mu)}$ are fixed. Here note that $\kappa$ and $\Omega_H^{(\mu)}$ are surface gravity and angular velocities, respectively, for $\phi_0$.

Up to this point we have not yet used explicitly the fact that $\tilde{\xi}_0^a = 0$ on $\Sigma$ for the perturbed solution, although we have used it implicitly. By using it explicitly, we can rewrite the l.h.s. of (2.36) in a useful form. The result is

$$\tilde{\delta}\int_\Sigma \mathbf{Q}[\tilde{\xi}_0] = \frac{1}{2\pi}\tilde{\delta}S, \tag{2.37}$$

where $S$ is defined by (2.30). (For explicit manipulations, see the proof of Theorem 6.1 of Iyer-Wald [48].)

Finally, we obtain the first law (2.29) for non-stationary perturbations $\tilde{\delta}\phi$ about a stationary black hole solution $\phi_0$.

Now we comment on entropy for the perturbed, non-stationary black hole.

As stated above, the $(n-2)$-surface $\Sigma$ has no meaning for the perturbed solution. (It is nothing but a surface on which $\tilde{\xi}_0^a$ vanishes.) In general, it does not lie on the event (or apparent) horizon for the perturbed solution. Hence, entropy evaluated on $\Sigma$ may not coincide with that on a cross section of the perturbed horizon, provided that the entropy is defined as $2\pi$ times an integral of $\mathbf{Q}[\tilde{\xi}_0]$ for both $(n-2)$-surfaces.

Note that it is in general impossible to make gauge transformation so that $\Sigma$ lie on a horizon cross section, if entropy (eg. a quarter of area in general relativity) on $\Sigma$ is different from entropy on a horizon cross section. The difference is given by $2\pi$ times an integral of the Noether current $\mathbf{j}[\tilde{\xi}_0]$ over a hypersurface whose boundary is a union of $\Sigma$ and a cross section of the perturbed horizon. Since it is natural to assign black hole entropy to the horizon cross section [48], it might be expected that there appears an extra term corresponding to the integral of $\mathbf{j}[\tilde{\xi}_0]$ in the first law.

However, as shown in the next paragraph, the integral of $\mathbf{j}[\tilde{\xi}_0]$ vanishes to first order in $\tilde{\delta}\phi$ [1]. Thus, $\tilde{\delta}S$ evaluated on $\Sigma$ gives the correct variation of entropy defined on the horizon to first order in $\tilde{\delta}\phi$. This means that the extra term does not appear and that the first law of Ref. [48] derived in this section for non-stationary perturbation about a stationary black hole is the correct formula.

Let us show the above statement. Since $\tilde{\xi}^a_0 = 0$ on $\Sigma$ and $\mathcal{L}_{\tilde{\xi}_0}\phi_0 = 0$, the Noether current $\mathbf{j}[\tilde{\xi}_0]$ vanishes on $\Sigma$ for the unperturbed solution by the definition (2.7). Hence, for the perturbed solution, the Noether current is at least first order in $\tilde{\delta}\phi$ on $\Sigma$. On the other hand, deviation of a horizon cross section from $\Sigma$ is at least first order. Therefore, the integral of $\mathbf{j}[\tilde{\xi}_0]$ over a hypersurface connecting $\Sigma$ and the perturbed horizon cross section is at least second order in $\tilde{\delta}\phi$.

Finally, let us apply the first law of this subsection to a stationary perturbation. The result is the same as that derived in the previous subsection. It is evident that the gauge condition used in this subsection is weaker than that used in the previous subsection. In fact, $\tilde{\xi}^a$ ($\neq \tilde{\xi}^a_0$ for a perturbed solution) is not fixed in the former condition. Hence, it can be concluded that the minimal set of gauge conditions necessary for the derivation of the first law is that $t^a$ and $\varphi^a_{(\mu)}$ are fixed at spatial infinity.

## 2.2 The first law of black hole dynamics

As will be seen in section 2.3, the first law of black hole statics is used in a (quasi-stationary but) dynamical situation to prove the generalized second law [52, 10], which is a natural generalization of both the second law (or area law [12]) of black hole and the second law of usual thermodynamics. In the proof, by assuming quasi-stationarity, the use of the first law of statics can be justified to relate a small change from an initial stationary black hole to final stationary one. However, if we intend to extend the proof of the generalized second law to finite changes between two stationary black holes or a purely dynamical situation, the first law of black hole statics can not be used.

So, we want to extend the first law to a dynamical situation and call it a first law of black hole dynamics (BHD). It will be discussed in subsection 2.3.3 that the generalized second law might be extended to not quasi-stationary situations by using the first law of black hole dynamics.

In this section, for simplicity, we consider general relativity only. In subsection 2.2.1 we consider two non-statistical definitions of entropy for dynamic (non-stationary) black holes in spherical symmetry. The first is analogous to the original Clausius definition of thermodynamic entropy: there is a first law containing an energy-supply term which equals surface gravity times a total differential. The second is Wald's Noether-charge method, adapted to dynamic black holes by using the Kodama flow. Both definitions give the same answer for Einstein gravity: one-quarter the area of the trapping horizon [41]. In subsection 2.2.2, the first law of BHD is derived without assuming any symmetry and any asymptotic conditions [42].

[1] The author thanks Professor R. M. Wald for helpful comments on this point.

In the derivation, a definition of dynamical surface gravity is proposed.

### 2.2.1 Dynamic black hole entropy in general relativity with spherical symmetry

It is generally thought that black-hole entropy should have a statistical origin, presumably in a quantum theory of gravity. This is, of course, due to the definition of entropy in statistical mechanics. However, it should be remembered that the original concept of entropy was not statistical [53]. The original argument of Clausius was that, in a cyclic reversible process, the total heat supply $\delta Q$ divided by temperature $\vartheta$ should vanish. Thus in any reversible process, $\delta Q/\vartheta$ should be the total differential $dS$ of a state function $S$, the entropy. Moreover, in irreversible processes, there should be a second law $dS \geq \delta Q/\vartheta$. The heat supply also occurs in a first law $dU = \delta Q + \delta W$, where $U$ is the internal energy and $\delta W$ the work being done. These are basic laws of thermodynamics as stated in typical textbooks and originally formulated by Clausius before the invention of statistical mechanics. In this subsection, we argue that there is a similar concept of entropy for dynamic black holes, suggested by the mathematical structure of the first law: it contains an energy-supply term which equals surface gravity times a total differential.

#### Kodama vector and the first law of black hole dynamics in spherical symmetry

The relevant quantities and equations in spherical symmetry may be summarized as follows. The area $A$ or areal radius $r = \sqrt{A/4\pi}$ of the spheres of symmetry determines the 1-form

$$k = *dr \tag{2.38}$$

where $d$ is the exterior derivative and $*$ is the Hodge operator of the two-dimensional space normal to the spheres of symmetry. Henceforth, 1-forms and their vector duals with respect to the space-time metric will not be distinguished. Then $k$ is the divergence-free vector introduced by Kodama [54], which generates a preferred flow of time and is a dynamic analogue of a stationary Killing vector [55]. The active gravitational energy or mass is

$$E = (1 - dr \cdot dr)r/2 \tag{2.39}$$

where the dot denotes contraction. Misner and Sharp [56] originally defined $E$ and, thus, is called the Misner-Sharp energy. (Ref. [57] described its physical properties.) The dynamic surface gravity

$$\kappa = *dk/2 \tag{2.40}$$

was defined in Ref.[55] by analogy with the standard definition of stationary surface gravity. This reference also introduced two invariants of the energy tensor $T$: the energy density (work density)

$$w = -\mathrm{tr}\, T/2 \tag{2.41}$$

and the energy flux (localized Bondi flux)

$$\psi = T \cdot dr + w dr \tag{2.42}$$

where tr denotes the the two-dimensional normal trace. One may say that $(A, k)$ are the basic kinematic quantities, $(E, \kappa)$ the gravitational quantities and $(w, \psi)$ the relevant matter quantities. Instead of $\psi$ one may also use the divergence-free energy-momentum vector [55, 54, 57]

$$j = *\psi + wk. \tag{2.43}$$

Finally, the relevant components of the Einstein equation are [55]

$$E = r^2\kappa + 4\pi r^3 w \tag{2.44}$$

and[57]

$$Aj = *dE. \tag{2.45}$$

The latter may be rewritten as

$$dE = A\psi + wdV \tag{2.46}$$

where $V = \frac{4}{3}\pi r^3$ is the areal volume. This is the unified first law of Ref.[55]. One may regard $wdV$ as a type of work and $A\psi$ as an energy supply, analogous to heat supply $\delta Q = \oint q$, where $q$ is the heat flux. The energy supply can be written as

$$A\psi = \frac{\kappa dA}{8\pi} + rd\left(\frac{E}{r}\right). \tag{2.47}$$

The second term vanishes when projected along a black-hole horizon, defined as in Refs.[55, 58, 57] by a trapping horizon: a hypersurface where $dr$ is null, so that $E = r/2$. This also occurs for any hypersurface on which $E/r$ is constant, thereby covering any smooth space-time. The key point is that the first term is the product of surface gravity $\kappa$ and a total differential. Identifying $\kappa/2\pi$ as a temperature, this total differential therefore determines a *Clausius entropy* $A/4$. Note that this stems from a purely mathematical property of the energy supply occurring in the first law. The restriction to spherical symmetry will be removed in subsection 2.2.2.

**Wald-Kodama entropy**

Wald [33] also gave a definition of entropy

$$\kappa S = 2\pi \oint Q[\xi] \tag{2.48}$$

where $Q$ henceforth denotes a Noether charge 2-form obtained from a certain type of Lagrangian. For Einstein gravity, $Q_{ab} = -\epsilon_{abcd}\nabla^c\xi^d/16\pi$ [48], where $\epsilon$ is the space-time volume form, $\xi$ is a generating vector for the diffeomorphisms which is taken to be the horizon generating Killing vector of a stationary black hole, and $\kappa$ is the surface gravity corresponding to $\xi$. In spherical symmetry, we propose using the Kodama vector $k$ for $\xi$ to give an alternative definition of the entropy of dynamic black holes. We call it *Wald-Kodama entropy*. This prescription effectively corresponds to also replacing $\xi$ by $k$ in $S_2$ of Jacobsen et al. [49]. Then from the above expression for $Q$ by Wald's method,

$$\oint Q[k] = \frac{A\kappa}{8\pi}, \tag{2.49}$$

where the integral is over a sphere of symmetry. Thus the Wald-Kodama entropy is

$$S = A/4. \tag{2.50}$$

**Agreement and speculations**

So, the Wald-Kodama entropy agrees with the Clausius entropy. The motivations also seem similar, since Wald's construction involved a first law of black-hole statics based on perturbations $\delta$ of a stationary solution. This agreement suggests that some combination of the two methods may be useful in general, assuming neither stationarity nor Einstein gravity.

It is also interesting that both methods formally hold not just on a black-hole horizon, but anywhere in the space-time, as do all the equations displayed above. Whether any surface in any space-time should have an entropy related to its area is arguable, but this does concur with the entanglement entropy approach, which is discussed in section 3.3.

### 2.2.2 Qusi-local first law of black hole dynamics in general relativity

#### A general definition of a black hole

Here we would like to treat a dynamical, not necessarily asymptotically flat space-time. Even for such a general situation, there is a definition of a black hole. Namely, a *black hole* is defined as a future outer trapping horizon [58]. The *future outer trapping horizon* is the closure of a three-surface foliated by marginal surfaces on which $\theta_- < 0$ and $\mathcal{L}_-\theta_+ < 0$, where the *marginal surface* is a spatial two-surface on which one of two null expansions (which we have denoted by $\theta_+$) vanishes. Here $\theta_-$ is another null expansion and $\mathcal{L}_-$ is a Lie derivative w.r.t. a null vector defined below. For the purpose of this subsection, we only need the fact that $\theta_+\theta_- = 0$ along the horizon. Hence, the first law we shall obtain remains to hold for a general trapping horizon, i.e. a hypersurface foliated by marginal surfaces. We mention here that a trapping horizon can be regarded as a black hole, a white hole and a wormhole when it is future outer, past outer and temporal outer, respectively [59].

#### The double-null formalism

To investigate behavior of the trapping horizon, the so-called double-null formalism, or the $(2+2)$ decomposition, of general relativity is useful. Among several $(2+2)$-formalisms [60, 61], we adopt one based on Lie derivatives w.r.t null vectors developed by Hayward [61]. Let us review basic ingredients of the formalism. Suppose that a four-dimensional spacetime manifold $(M, g)$ is foliated (at least locally) by two families of null hypersurfaces $\Sigma^\pm$, each of which is parameterized by a scalar $\xi^\pm$, respectively. The null character is described by $g^{-1}(n^\pm, n^\pm) = 0$, where $n^\pm = -d\xi^\pm$ are normal 1-forms to $\Sigma^\pm$. The relative normalization of the null normals defines a function $f$ as $g^{-1}(n^+, n^-) = -e^f$. The intersections of $\Sigma^+(\xi^+)$ and $\Sigma^-(\xi^-)$ define a two-parameter family of two-dimensional spacelike surfaces $S(\xi^+, \xi^-)$. Hence, by introducing an intrinsic coordinate system $(\theta^1,\theta^2)$ of the 2-surfaces, the foliation is described by the imbedding $x = x(\xi^+, \xi^-; \theta^1, \theta^2)$.

For the imbedding, the intrinsic metric on the 2-surfaces is found to be $h = g + e^{-f}(n^+ \otimes n^- + n^- \otimes n^+)$. Correspondingly, the vectors $u_\pm = \partial/\partial\xi^\pm$ have 'shift vectors' $s_\pm = \perp u_\pm$, where $\perp$ indicates projection by $h$. The 4-dimensional metric is written in terms of $(h,f,s_\pm)$ as

$$g = \begin{pmatrix} h(s_+, s_+) & h(s_+, s_-) - e^{-f} & h(s_+) \\ h(s_-, s_+) - e^{-f} & h(s_-, s_-) & h(s_-) \\ h(s_+) & h(s_-) & h \end{pmatrix}. \tag{2.51}$$

Geometrical quantities such as *expansions* $\theta_\pm$, *shears* $\sigma_\pm$ and the *twist* $\omega$ are defined by $\theta_\pm = *\mathcal{L}_\pm * 1$, $\sigma_\pm = \perp \mathcal{L}_\pm h - \theta_\pm h$ and $\omega = e^f[l_-, l_+]/2$, where $*$ denotes the Hodge-dual operator of $h$, $l_\pm = u_\pm - s_\pm = e^{-f} g^{-1}(n^\mp)$ are null normal vectors to $\Sigma^\pm$, and $\mathcal{L}_\pm$ denotes the Lie derivative along $l_\pm$, respectively. It is possible to write down the Einstein tensor in terms of these geometrical quantities. The component

useful for our purpose is $G_{+-} = G(l_+, l_-)$, which is given by

$$2e^f G_{+-} = {}^{(2)}R + e^f(\mathcal{L}_+\theta_- + \mathcal{L}_-\theta_+ + 2\theta_+\theta_-) - 2\left[h(\omega,\omega) + \frac{1}{4}h^\sharp(df,df)\right] + \mathcal{D}^2 f. \tag{2.52}$$

Here $h^\sharp = g^{-1}hg^{-1}$ is $h$ raised by $g^{-1}$, $\mathcal{D}^2$ and ${}^{(2)}R$ are the two-dimensional Laplacian and the Ricci scalar both associated with the metric $h$.

### Hawking energy

Before deriving the first law we have to define energy and surface gravity in a quasi-local way. In spherical symmetry there is a widely accepted energy: the Misner-Sharp (MS) energy (2.39) [56]. In the previous subsection the MS energy is used to derive the first law of BHD in spherical symmetry [55]. In this subsection we adopt the Hawking energy [62], which reduces to the MS energy in spherical symmetry. It is defined by

$$E(\xi^+,\xi^-) = \frac{r}{16\pi}\int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}\left[{}^{(2)}R + e^f\theta_+\theta_-\right], \tag{2.53}$$

where $h$ is the determinant of the two-dimensional metric $h_{ab}$ and the area radius $r$ is defined by

$$r = \sqrt{A/4\pi}, A = \int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}. \tag{2.54}$$

### A proposal of dynamical surface gravity

In Ref. [55], a definition of dynamical surface gravity was proposed in spherical symmetry as (2.40). A natural generalization to a non spherically-symmetric case is

$$\kappa(\xi^+,\xi^-) = \frac{-1}{16\pi r}\int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}\; e^f(\mathcal{L}_+\theta_- + \mathcal{L}_-\theta_+ + \theta_+\theta_-). \tag{2.55}$$

This is the most simple generalization in the sense that it includes neither the shear $\sigma_{Aab}$ nor the twist $\omega^a$.

Note that this definition of surface gravity and the definition (2.53) of the Hawking energy are both quasi-local in the sense originally introduced by Penrose [63]. Hence we call the corresponding first law, which we shall derive below, *the quasi-local first law* of BHD.

### The dynamical first law

We now derive the quasi-local first law of BHD for the Hawking energy (2.53) and the surface gravity defined by (2.55). It is easy to show that

$$dE - \frac{\kappa}{8\pi}dA = wAdr + rd\left(\frac{E}{r}\right), \tag{2.56}$$

where $w$ is defined by

$$w = \frac{1}{A}\left(\frac{E}{r} - \kappa r\right). \tag{2.57}$$

Here note that '$d$' in Eq. (2.56) is not a variation in a space of stationary solutions of the Einstein equation as in the first law of black hole statistics, but is the differentiation w.r.t. the parameters $\xi^\pm$ of the spacetime foliation. (For example, $dE = d\xi^+\partial_+E + d\xi^-\partial_-E$.) We mention that Eq. (2.56) holds independently of the definitions of $E$ and $\kappa$ while the following arguments depend on the definitions.

Since the Gauss-Bonnet theorem says that

$$\int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}\ {}^{(2)}R = 8\pi(1-\gamma),$$

where $\gamma$ is the genus or number of handles of $S(\xi^+,\xi^-)$, the energy divided by area radius is given by $E/r = (1-\gamma)/2$ on a marginal surface and is a constant. Thus,

$$E' = \frac{\kappa}{8\pi}A' + wAr', \tag{2.58}$$

where the prime denotes the derivative along the trapping horizon. This is the quasi-local first law of BHD. Note that this also holds along any hypersurface foliated by 2-surfaces on which $E/r$ is constant.

**The work term**

By using Eq. (2.52) it is easy to show that $w$ is written as follows.

$$w = \rho_m + \rho_j, \tag{2.59}$$

where the averaged matter energy density $\rho_m$ and the effective angular energy density $\rho_j$ are defined by

$$\begin{aligned} \rho_m &= \frac{1}{8\pi A}\int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}e^f G_{+-}, \\ \rho_j &= \frac{1}{8\pi A}\int_{S(\xi^+,\xi^-)} d^2\theta\sqrt{h}\left[h(\omega,\omega) + \frac{1}{4}h^\sharp(df,df)\right]. \end{aligned} \tag{2.60}$$

The Einstein equation $G = 8\pi T$ says that $\rho_m$ is $e^f T(l_+,l_-)$ averaged over the 2-surface. It seems that $\rho_j$ represents effective energy density due to angular momentum.

The term $wAr'$ should be a work term done along the horizon. For example, for an electromagnetic field, the term $\rho_m Ar'$ reduces to the electromagnetic work done along the horizon [55]. It seems that the term $\rho_j Ar'$ is a work associated with angular momentum of the trapping horizon.

**Comments**

In this subsection the quasi-local first law of black hole dynamics has been derived without assuming any symmetry and any asymptotic condition. In the derivation we have given a new definition of dynamical surface gravity. In spherical symmetry it reduces to that defined in Ref. [55].

By using the quasi-local first law derived in this subsection, it might be possible to extend a proof of the generalized second law to not quasi-stationary situations. (See subsection 2.3.3.)

Besides the first law derived in this subsection, there exist the second law [58] and the third law [14] for the trapping horizon (or apparent horizon). It seems that by using these laws we can formulate black hole thermodynamics consistently as trapping horizon dynamics. However, for this purpose, there is an important open question: we have to associate temperature of quantum fields with the trapping horizon. All we can say here is that the temperature may be given by $\hbar\kappa/2\pi$, where $\kappa$ is the surface gravity introduced here.

The final comment is in order. The surface gravity $\kappa(\xi^+,\xi^-)$ is an invariant of a double-null foliation at the surface. Since a non-null trapping horizon locally determines a unique double-null foliation, the surface gravity is definitely an invariant

of the trapping horizon if the horizon is not null. On the other hand, the null case is ambiguous because of the freedom to rescale the other null direction. Fixing this would require some kind of limiting argument that might be effectively a zeroth law. Therefore, we have to impose an auxiliary condition for the surface gravity $\kappa(\xi^+, \xi^-)$ to work well when the trapping horizon is null. Since surface gravity seems to be related to temperature of quantum fields as stated above, it will be valuable to investigate the auxiliary condition in detail [42].

## 2.3 The generalized second law

The generalized second law of black hole thermodynamics is a statement that a sum of black hole entropy and thermodynamic entropy of matter fields outside the horizon does not decrease [64, 65, 52, 10], where the black hole entropy is defined as a quarter of the area of the horizon. Namely it says that an entropy of the whole system does not decrease. It interests us in a quite physical sense since it links a world inside a black hole and our thermodynamic world. In particular it gives a physical meaning to black hole entropy indirectly since it concerns the sum of black hole entropy and ordinary thermodynamic entropy, and since physical meaning of the latter is well-known by statistical mechanics.

Frolov and Page [52] proved the generalized second law for a quasi-stationary eternal black hole by assuming that a state of matter fields on the past horizon is thermal one and that a set of radiation modes on the past horizon and a set of radiation modes on the past null infinity are quantum mechanically uncorrelated. The assumption is reasonable for the eternal case since a black hole emit a thermal radiation (the Hawking radiation). When we attempt to apply their proof to a non-eternal black hole which is formed by gravitational collapse, we might expect that things would go well by simply replacing the past horizon with a null surface at a moment of a formation of a horizon ($v = v_0$ surface in *Figure* 2.1). However, it is not the case since the above assumption does not hold in this case. The reason is that on a background describing gravitational collapse the thermal radiation is observed not at the moment of the horizon formation but at the future null infinity and that any modes on the future null infinity have correlation with modes on the past null infinity located after the horizon formation. The correlation can be seen in the equation (2.65) of this section explicitly. Thus, their proof does not hold for the case in which a black hole is formed by gravitational collapse. Since a black hole is thought to be formed by gravitational collapse in astrophysical situation, we want to prove the generalized second law in this case.

The rest of this section is organized as follows. In subsection 2.3.1 we consider a real massless scalar field in a background of a gravitational collapse to show that a thermal state with special values of temperature and chemical potential evolves to a thermal state with the same temperature and the same chemical potential. These special values are determined by the background geometry. In subsection 2.3.2, first, the generalized second law is rewritten as an inequality which states that there is a non-decreasing functional of a density matrix of matter fields. After that, we give a theorem which shows an inequality between functionals of density matrices. Finally, we apply it to the scalar field investigated in subsection 2.3.1 to prove the generalized second law for the quasi-stationary background. In subsection 2.3.3 we summarize this section.

### 2.3.1 A massless scalar field in black hole background

In this subsection we consider a real massless scalar field in a curved background which describes formation of a quasi-stationary black hole. Let us denote the past

null infinity by $\mathcal{I}^-$, the future null infinity by $\mathcal{I}^+$ and the future event horizon by $H^+$. Introduce the usual null coordinates $u$ and $v$, and suppose that the formation of the event horizon $H^+$ is at $v = v_0$ (see *Figure* 2.1). On $\mathcal{I}^-$ and $\mathcal{I}^+$, by virtue of the asymptotic flatness, there is a natural definition of Hilbert spaces $\mathcal{H}_{\mathcal{I}^-}$ and $\mathcal{H}_{\mathcal{I}^+}$ of mode functions with positive frequencies [67]. The Hilbert spaces $\mathcal{F}(\mathcal{H}_{\mathcal{I}^\pm})$ of all asymptotic states are defined as follows with a suitable completion (symmetric Fock spaces):

$$\mathcal{F}(\mathcal{H}_{\mathcal{I}^\pm}) \equiv \boldsymbol{C} \oplus \mathcal{H}_{\mathcal{I}^\pm} \oplus (\mathcal{H}_{\mathcal{I}^\pm} \otimes \mathcal{H}_{\mathcal{I}^\pm})_{sym} \oplus \cdots,$$

where $(\cdots)_{sym}$ denotes the symmetrization ($(\xi \otimes \eta)_{sym} = \frac{1}{2}(\xi \otimes \eta + \eta \otimes \xi)$, etc.). Physically, $\boldsymbol{C}$ denotes the vacuum state, $\mathcal{H}_{\mathcal{I}^\pm}$ one particle states, $(\mathcal{H}_{\mathcal{I}^\pm} \otimes \mathcal{H}_{\mathcal{I}^\pm})_{sym}$ two particle states, etc.. We suppose that all our observables are operators on $\mathcal{F}(\mathcal{H}_{\mathcal{I}^\pm})$ since we observe a radiation of the scalar field radiated by the black hole at places far away from it. In this sense $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ are quite physical. Next let us consider how to set an initial state of the scalar field. We want to see a response of the scalar field on the quasi-stationary black hole background which is formed by gravitational collapse of other materials (a dust, a fluid, etc.). Hence as the initial state at $\mathcal{I}^-$ we consider a state such that it includes no excitations of modes located before the formation of the horizon (no excitation at $v < v_0$). A space of all such states is a subspace of $\mathcal{F}(\mathcal{H}_{\mathcal{I}^-})$, and we denote it by $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$. We like to derive a thermal property of a scattering process of the scalar field by the quasi-stationary black hole. Hence we consider density matrices on $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ and $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$. Denote a space of all density matrices on $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ by $\mathcal{P}$ and a space of all density matrices on $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ by $\tilde{\mathcal{P}}$.

Let us discuss the evolution of a state at $\mathcal{I}^-$ to future. Since $\mathcal{I}^+$ is not a Cauchy surface because of the existence of $H^+$, $\mathcal{F}(\mathcal{H}_{\mathcal{I}^-})$ is mapped not to $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ but to $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+}) \otimes \mathcal{F}(\mathcal{H}_{H^+})$ by a unitary evolution, where $\mathcal{H}_{H^+}$ is a Hilbert space of mode functions on the horizon with a positive frequency, and $\mathcal{F}(\mathcal{H}_{H^+})$ is a Hilbert space of all states on $H^+$ defined as a symmetric Fock space (see the definition of $\mathcal{F}(\mathcal{H}_{\mathcal{I}^\pm})$). Although there is no natural principle to determine positive frequency modes (equivalently, there is no natural definition of the particle concept) on $H^+$, how to define $\mathcal{H}_{H^+}$ does not affect the result since we shall trace out the degrees of freedom of $\mathcal{F}(\mathcal{H}_{H^+})$ (see (2.61)). To describe the evolution of a quantum state of the scalar field from $\mathcal{F}(\mathcal{H}_{\mathcal{I}^-})$ to $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+}) \otimes \mathcal{F}(\mathcal{H}_{H^+})$ an S-matrix is introduced [67]. For a given initial state $|\psi\rangle$ in $\mathcal{F}(\mathcal{H}_{\mathcal{I}^-})$, the corresponding final state in $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+}) \otimes \mathcal{F}(\mathcal{H}_{H^+})$ is $S|\psi\rangle$. Then the corresponding evolution from $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ to $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ is obtained by restricting $S$ to $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$, and we denote it by $S$, too. The S-matrix elements was given by Wald [67].

### Superscattering matrix $T$

Suppose that the initial state of the scalar field is $|\phi\rangle$ ($\in \mathcal{F}(\mathcal{H}_{\mathcal{I}^-})$) and that the corresponding final state is observed at $\mathcal{I}^+$ (see the argument after the definition of $\mathcal{F}(\mathcal{H}_{\mathcal{I}^\pm})$). Formally the observation corresponds to a calculation of a matrix element $\langle\phi|S^\dagger O S|\phi\rangle$, where $S$ is the S-matrix which describes the evolution of the scalar field from $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ to $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+}) \otimes \mathcal{F}(\mathcal{H}_{H^+})$ and $O$ is a self-adjoint operator on $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ corresponding to a quantity we want to observe. The matrix element can be rewritten in the following convenient fashion:

$$\langle\phi|S^\dagger O S|\phi\rangle = \boldsymbol{Tr}_{\mathcal{I}^+}\left[O\rho_{red}\right],$$

where

$$\rho_{red} = \boldsymbol{Tr}_{H^+}\left[S|\phi\rangle\langle\phi|S^\dagger\right],$$

$\boldsymbol{Tr}_{\mathcal{I}^+}$ and $\boldsymbol{Tr}_{H^+}$ denote partial trace over $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ and $\mathcal{F}(\mathcal{H}_{H^+})$, respectively. In viewing this expression we are lead to an interpretation that the corresponding final

state at $\mathcal{I}^+$ is represented by the reduced density matrix $\rho_{red}$. Next we generalize this argument to a wider class of initial states, which includes all mixed states. For this case an initial state is represented not by an element of $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ but by an element of $\mathcal{P}$ (a density matrix on $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$). Its evolution to $\mathcal{I}^+$ is represented by the so-called superscattering matrix $T$ defined as follows: let $\rho$ ($\in \mathcal{P}$) be an initial density matrix then the corresponding final density matrix $T(\rho)$ ($\in \tilde{\mathcal{P}}$) is

$$T(\rho) = \boldsymbol{Tr}_{H^+}\left[S\rho S^\dagger\right]. \tag{2.61}$$

Note that $T$ is a linear map from $\mathcal{P}$ into $\tilde{\mathcal{P}}$.

**Thermodynamic property of $T$**

Let us calculate a conditional probability defined as follows:

$$P(\{n_{i\rho}\}|\{n_{i\gamma}\}) \equiv \langle\{n_{i\rho}\}|T\left(|\{n_{i\gamma}\}\rangle\langle\{n_{i\gamma}\}|\right)|\{n_{i\rho}\}\rangle, \tag{2.62}$$

where

$$\begin{aligned} |\{n_{i\gamma}\}\rangle &\equiv \left[\prod_i \frac{1}{\sqrt{n_{i\gamma}!}}\left(a^\dagger(A\ {}_i\gamma)\right)^{n_{i\gamma}}\right]|0\rangle, \\ |\{n_{i\rho}\}\rangle &\equiv \left[\prod_i \frac{1}{\sqrt{n_{i\rho}!}}\left(a^\dagger({}_i\rho)\right)^{n_{i\rho}}\right]|0\rangle. \end{aligned} \tag{2.63}$$

$|\{n_{i\gamma}\}\rangle$ is a state in $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$ characterized by a set of integers $n_{i\gamma}$ $(i = 1, 2, \cdots)$ and $|\{n_{i\rho}\}\rangle$ is a state in $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$ characterized by a set of integers $n_{i\rho}$ $(i = 1, 2, \cdots)$. Therefore $P(\{n_{i\rho}\}|\{n_{i\gamma}\})$ is a conditional probability for the final state to be $|\{n_{i\rho}\}\rangle$ when the initial state is specified to be $|\{n_{i\gamma}\}\rangle$. In the expressions, $A$ is a part of a representation of a Bogoliubov transformation [67], which represents a map from $\mathcal{H}_{\mathcal{I}^+} \oplus \mathcal{H}_{H^+}$ to $\mathcal{H}_{\mathcal{I}^-}$, and ${}_i\gamma$ is a unit vector in $\mathcal{H}_{\mathcal{I}^+} \oplus \mathcal{H}_{H^+}$ such that $A\ {}_i\gamma$ corresponds to a wave packet whose peak is located at a point on $\mathcal{I}^-$ later than the formation of the horizon ($v > v_0$) [67]. On the other hand, ${}_i\rho$ is a unit vector in $\mathcal{H}_{\mathcal{I}^+}$ and corresponds to a wave packet on $\mathcal{I}^+$ [67] (see *Figure* 2.1). The probability (2.62) is a generalization of $P(k|j)$ investigated by Panangaden and Wald [9]. (Our $P(\{n_{i\rho}\}|\{n_{i\gamma}\})$ reduces to $P(k|j)$ of Panangaden-Wald when $n_{i_0\gamma} = j$, $n_{i_0\rho} = k$ and $n_{i\gamma} = n_{i\rho} = 0$ for all $i$ other than $i_0$. Here $i_0$ is an arbitrary fixed value of $i$.) Evidently, our conditional probability $P(\{n_{i\rho}\}|\{n_{i\gamma}\})$ includes more abundant information [2] about a response of the scalar field than $P(k|j)$. In fact, any initial states on $\mathcal{I}^-$, which include no excitation before the formation of the horizon ($v < v_0$.), can be represented by using the basis $\{|\{n_{i\gamma}\}\rangle\}$ and any final states on $\mathcal{I}^+$ can be expressed by the basis $\{|\{n_{i\rho}\}\rangle\}$, i.e. a set of all $|\{n_{i\gamma}\}\rangle$ generates $\mathcal{F}(\mathcal{H}_{\mathcal{I}^-(v>v_0)})$ and a set of all $|\{n_{i\rho}\}\rangle$ generates $\mathcal{F}(\mathcal{H}_{\mathcal{I}^+})$. This is the very reason why we have generalized $P(k|j)$ to $P(\{n_{i\rho}\}|\{n_{i\gamma}\})$.

By using the S-matrix elements given in [67], the conditional probability is rewritten as follows (see appendix A.1 for its derivation):

$$\begin{aligned} &P(\{n_{i\rho}\}|\{n_{i\gamma}\}) \\ &= \prod_i \Big[(1-x_i)x_i^{2n_{i\rho}}\left(1-|R_i|^2\right)^{n_{i\gamma}+n_{i\rho}} \\ &\times \sum_{l_i=0}^{\min(n_{i\gamma},n_{i\rho})} \sum_{m_i=0}^{\min(n_{i\gamma},n_{i\rho})} \frac{\left[-|R_i|^2/(1-|R_i|^2)\right]^{l_i+m_i} n_{i\gamma}!n_{i\rho}!}{l_i!(n_{i\gamma}-l_i)!(n_{i\rho}-l_i)!m_i!(n_{i\gamma}-m_i)!(n_{i\rho}-m_i)!} \end{aligned}$$

[2] All the information about the response of the scalar field is included in $T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}}$ defined in Lemma 2.

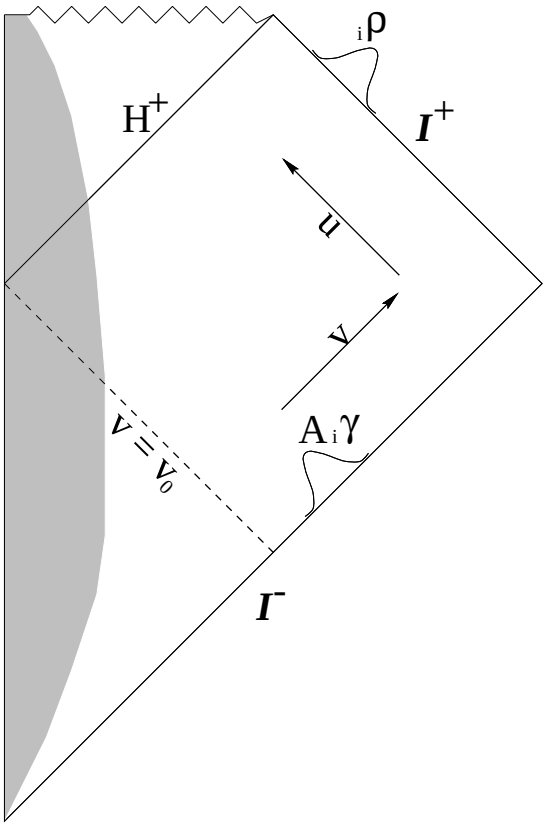


Figure 2.1: A conformal diagram of a background geometry which describes a gravitational collapse. $\mathcal{I}^-$ and $\mathcal{I}^+$ are the past null infinity and the future null infinity, respectively and $H^+$ is the future event horizon. Shaded region represents collapsing materials which forms the black hole. Besides the collapsing matter, we consider a real massless scalar field and investigate a scattering problem by the black hole after its formation ($v > v_0$). Thus we specify possible initial states at $\mathcal{I}^-$ to those states which are excited from the vacuum by only modes whose support is within $v > v_0$ (elements of $\mathcal{F}_{\mathcal{I}^-(v>v_0)}$), and possible mixed states constructed from them (elements of $\mathcal{P}$). In the diagram, $A\ {}_i\gamma$ $(i = 1, 2, \cdots)$ is a mode function corresponding to a wave packet whose peak is at $v > v_0$ on $\mathcal{I}^-$, ${}_i\rho$ $(i = 1, 2, \cdots)$ is a mode function corresponding to a wave packet on $\mathcal{I}^+$.

$$\times \sum_{n_i=n_{i\rho}-\min(l_i,m_i)}^{\infty} \frac{n_i!(n_i-n_{i\rho}+n_{i\gamma})!}{(n_i-n_{i\rho}+l_i)!(n_i-n_{i\rho}+m_i)!}(x_i^2|R_i|^2)^{n_i-n_{i\rho}}\Bigg], \quad (2.64)$$

where $R_i$ is a reflection coefficient for the mode specified by the integer $i$ on the Schwarzschild metric (see appendix A.1) and $x_i$ is a constant defined by $x_i = \exp(-\pi(\omega_i - \Omega_{BH}m_i)/\kappa)$. In the expression, $\omega_i$ and $m_i$ are the frequency and the azimuthal angular momentum quantum number of the mode specified by the integer $i$, $\Omega_{BH}$ and $\kappa$ are the angular velocity and the surface gravity of the black hole.

Now, the expression in the squared bracket in (2.64) appears also in the calculation of $P(k|j)$. Using the result of [9], it is easily shown that

$$P(\{n_{i\rho}\}|\{n_{i\gamma}\}) = \prod_i \left[ K_i \sum_{s_i=0}^{\min(n_{i\rho},n_{i\gamma})} \frac{(n_{i\rho}+n_{i\gamma}-s_i)!v_i^{s_i}}{s_i!(n_{i\rho}-s_i)!(n_{i\gamma}-s_i)!} \right], \quad (2.65)$$

where

$$\begin{aligned} K_i &= \frac{(1-x_i)x_i^{2n_{i\rho}}\left(1-|R_i|^2\right)^{n_{i\gamma}+n_{i\rho}}}{(1-|R_i|^2x_i^2)^{n_{i\gamma}+n_{i\rho}+1}}, \\ v_i &= \frac{\left(|R_i|^2-x_i^2\right)\left(1-|R_i|^2x_i^2\right)}{\left(1-|R_i|^2\right)^2x_i^2}. \end{aligned}$$

This is a generalization of the result of [9], and the following lemma is easily derived by using this expression.

**Lemma 1** *For the conditional probability defined by (2.62) the following equality holds:*

$$\begin{aligned} &P(\{n_{i\rho}=k_i\}|\{n_{i\gamma}=j_i\})e^{-\beta_{BH}\sum_i j_i(\omega_i-\Omega_{BH}m_i)} \\ &= P(\{n_{i\rho}=j_i\}|\{n_{i\gamma}=k_i\})e^{-\beta_{BH}\sum_i k_i(\omega_i-\Omega_{BH}m_i)}, \end{aligned} \quad (2.66)$$

*where $\omega_i$ and $m_i$ are the frequency and the azimuthal angular momentum quantum number of the mode specified by $i$, $\Omega_{BH}$ is the angular velocity of the horizon and*

$$\beta_{BH} \equiv 2\pi/\kappa.$$

*Here $\kappa$ is the surface gravity of the black hole.*

Note that $\beta_{BH}^{-1}$ is the Hawking temperature of the black hole. This lemma states that a detailed balance condition holds [3]. Summing up about all $k$'s, we expect that a thermal density matrix $\rho_{th}(\beta_{BH},\Omega_{BH})$ in $\mathcal{P}$ with a temperature $\beta_{BH}^{-1}$ and a chemical potential $\Omega_{BH}$ for azimuthal angular momentum quantum number will be mapped by $T$ to a thermal density matrix $\tilde{\rho}_{th}(\beta_{BH},\Omega_{BH})$ in $\tilde{\mathcal{P}}$ with the same temperature and the same chemical potential. To show that this expectation is true, we have to prove that all off-diagonal elements of $T\left(\rho_{th}(\beta_{BH},\Omega_{BH})\right)$ are zero. For this purpose the following lemma is proved in Appendix A.2.

**Lemma 2** *Denote a matrix element of $T$ as*

$$T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}} \equiv \langle\{n_{i\rho}\}|T\left(|\{n_{i\gamma}\}\rangle\langle\{n'_{i\gamma}\}|\right)|\{n'_{i\rho}\}\rangle. \quad (2.67)$$

[3] It guarantees that a thermal distribution of any temperature is mapped to a thermal distribution of some other temperature closer to the Hawking temperature, as far as the diagonal elements are concerned.

*Then*

$$T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}} = 0, \tag{2.68}$$

*unless*

$$n_{i\gamma} - n'_{i\gamma} = n_{i\rho} - n'_{i\rho} \tag{2.69}$$

*for* $\forall i$.

Lemma 2 shows that all off-diagonal elements of $T(\rho)$ in the basis $\{|\{n_{i\rho}\}\rangle\}$ vanish if all off-diagonal elements of $\rho$ in the basis $\{|\{n_{i\gamma}\}\rangle\}$ is zero. Thus, combining it with Lemma 1 and the well-known fact that $|0\rangle$ is mapped to the thermal state, the following theorem is easily proved. Note that a set of all $|\{n_{i\gamma}\}\rangle\langle\{n'_{i\gamma}\}|$ generates $\mathcal{P}$ and a set of all $|\{n_{i\rho}\}\rangle\langle\{n'_{i\rho}\}|$ generates $\tilde{\mathcal{P}}$ (see the argument below (2.63)).

**Theorem 3** *Consider the linear map $T$ defined by (2.61) for a real, massless scalar field on a background geometry which describes a formation of a quasi-stationary black hole. Then*

$$T\left(\rho_{th}(\beta_{BH}, \Omega_{BH})\right) = \tilde{\rho}_{th}(\beta_{BH}, \Omega_{BH}), \tag{2.70}$$

*where*

$$\begin{aligned}
\rho_{th}(\beta_{BH}, \Omega_{BH}) &\equiv Z^{-1} \sum_{\{n_{i\gamma}\}} e^{-\beta_{BH} \sum_i n_{i\gamma}(\omega_i - \Omega_{BH} m_i)} |\{n_{i\gamma}\}\rangle\langle\{n_{i\gamma}\}|, \\
\tilde{\rho}_{th}(\beta_{BH}, \Omega_{BH}) &\equiv Z^{-1} \sum_{\{n_{i\rho}\}} e^{-\beta_{BH} \sum_i n_{i\rho}(\omega_i - \Omega_{BH} m_i)} |\{n_{i\rho}\}\rangle\langle\{n_{i\rho}\}|, \\
Z &\equiv \sum_{\{j_i\}} e^{-\beta_{BH} \sum_i j_i(\omega_i - \Omega_{BH} m_i)}.
\end{aligned} \tag{2.71}$$

$\rho_{th}(\beta_{BH}, \Omega_{BH})$ and $\tilde{\rho}_{th}(\beta_{BH}, \Omega_{BH})$ can be regarded as 'grand canonical ensemble' in $\mathcal{P}$ and $\tilde{\mathcal{P}}$ respectively, which have a common temperature $\beta_{BH}^{-1}$ and a common chemical potential $\Omega_{BH}$ for azimuthal angular momentum quantum number. Thus the theorem says that the 'grand canonical ensemble' at $\mathcal{I}^-$ $(v > v_0)$ with special values of temperature and chemical potential evolves to a 'grand canonical ensemble' at $\mathcal{I}^+$ with the same temperature and the same chemical potential. Note that the special values $\beta_{BH}^{-1}$ and $\Omega_{BH}$ are determined by the background geometry: $\beta_{BH}^{-1}$ is the Hawking temperature and $\Omega_{BH}$ is the angular velocity of the black hole formed. This result is used in subsection 2.3.2 to prove the generalized second law for the quasi-stationary black hole.

### 2.3.2 A proof of the generalized second law

The generalized second law of black hole thermodynamics is

$$\Delta S_{BH} + \Delta S_{matter} \geq 0, \tag{2.72}$$

where $\Delta$ denotes a change of quantities under an evolution of the system, $S_{BH}$ and $S_{matter}$ are black hole entropy of the black hole and thermodynamic entropy of the matter fields, respectively. For a quasi-stationary black hole, using the first law of the black hole thermodynamics

$$\Delta S_{BH} = \beta_{BH}(\Delta M_{BH} - \Omega_{BH} \Delta J_{BH}),$$

the conservation of total energy

$$\Delta M_{BH} + \Delta E_{matter} = 0$$

and the conservation of total angular momentum

$$\Delta J_{BH} + \Delta L_{matter} = 0,$$

it is easily shown that the generalized second law is equivalent to the following inequality:

$$\Delta S_{matter} - \beta_{BH}(\Delta E_{matter} - \Omega_{BH}\Delta L_{matter}) \geq 0, \tag{2.73}$$

where $\beta_{BH}$, $\Omega_{BH}$, $M_{BH}$ and $J_{BH}$ are the inverse temperature, the angular velocity, the mass and the angular momentum of the black hole; $E_{matter}$ and $L_{matter}$ are the energy and the azimuthal component of angular momentum of the matter fields. Thus this is of the form

$$U[\tilde{\rho}_0; \beta_{BH}, \Omega_{BH}) \geq U[\rho_0; \beta_{BH}, \Omega_{BH}), \tag{2.74}$$

where $U$ is a functional of a density matrix of the matter fields defined by

$$U[\rho; \beta_{BH}, \Omega_{BH}) \equiv -\boldsymbol{Tr}\left[\rho \ln \rho\right] - \beta_{BH}\left(\boldsymbol{Tr}[\boldsymbol{E}\rho] - \Omega_{BH}\boldsymbol{Tr}[\boldsymbol{L}_z\rho]\right), \tag{2.75}$$

$\rho_0$ and $\tilde{\rho}_0$ are an initial density matrix and the corresponding final density matrix respectively. In the expression $\boldsymbol{E}$ and $\boldsymbol{L}_z$ are operators corresponding to the energy and the azimuthal component of the angular momentum. Note that (2.74) is an inequality between functionals of a density matrix of matter fields [4]. We will prove the generalized second law by showing that this inequality holds. Actually we do it for a quasi-stationary black hole which is formed by gravitational collapse, using the results of subsection 2.3.1 and a theorem given in the following.

### Non-decreasing functional

In this subsection a theorem which makes it possible to construct a functional which does not decrease by a physical evolution. It is a generalization of a result of [66]. After that, we derive (2.74) for a quasi-stationary black hole which arises from gravitational collapse, applying the theorem to the scalar field investigated in subsection 2.3.1.

Let us consider Hilbert spaces $\mathcal{F}$ and $\tilde{\mathcal{F}}$. First we give some definitions needed for the theorem.

**Definition 4** *A linear bounded operator $\rho$ on $\mathcal{F}$ is called a density matrix, if it is self-adjoint, positive semi-definite and satisfies*

$$\boldsymbol{Tr}\rho = 1.$$

In the rest of this section we denote a space of all density matrices on $\mathcal{F}$ as $\mathcal{P}(\mathcal{F})$. Evidently $\mathcal{P}(\mathcal{F})$ is a linear convex set rather than a linear set.

**Definition 5** *A map $\mathcal{T}$ of $\mathcal{P}(\mathcal{F})$ into $\mathcal{P}(\tilde{\mathcal{F}})$ is called linear, if*

$$\mathcal{T}\left(a\rho_1 + (1-a)\rho_2\right) = a\mathcal{T}(\rho_1) + (1-a)\mathcal{T}(\rho_2)$$

*for $0 \leq {}^{\forall}a \leq 1$ and ${}^{\forall}\rho_1, {}^{\forall}\rho_2$ $(\in \mathcal{P}(\mathcal{F}))$.*

By this definition it is easily proved by induction that

$$\mathcal{T}\left(\sum_{i=1}^{N} a_i\rho_i\right) = \sum_{i=1}^{N} a_i\mathcal{T}(\rho_i), \tag{2.76}$$

if $a_i \geq 0$, $\sum_{i=1}^{N} a_i = 1$ and $\rho_i \in \mathcal{P}(\mathcal{F})$.

Now we prove the following lemma which concerns the $N \to \infty$ limit of the left hand side of (2.76). We use this lemma in the proof of theorem 7.

[4] Information about the background geometry appears in the inequality as the variables $\beta_{BH}$ and $\Omega_{BH}$ which parameterize the functional.

**Lemma 6** *Consider a linear map $\mathcal{T}$ of $\mathcal{P}(\mathcal{F})$ into $\mathcal{P}(\tilde{\mathcal{F}})$ and an element $\rho_0$ of $\mathcal{P}(\mathcal{F})$. For a diagonal decomposition*

$$\rho_0 = \sum_{i=1}^{\infty} p_i |i\rangle\langle i|,$$

*define a series of density matrices of the form*

$$\rho_n = \sum_{i=1}^{n} p_i/a_n |i\rangle\langle i| \qquad (n = N, N+1, \cdots), \tag{2.77}$$

*where*

$$a_n \equiv \sum_{i=1}^{n} p_i$$

*and $N$ is large enough that $a_N > 0$. Then*

$$\lim_{n\to\infty} \langle\Phi|\mathcal{T}(\rho_n)|\Psi\rangle = \langle\Phi|\mathcal{T}(\rho_0)|\Psi\rangle \tag{2.78}$$

*for arbitrary elements $|\Phi\rangle$ and $|\Psi\rangle$ of $\tilde{\mathcal{F}}$.*

This lemma says that $\mathcal{T}(\rho_n)$ has a weak-operator-topology-limit $\mathcal{T}(\rho_0)$.

**Proof**

By definition,

$$\rho_0 = a_n \rho_n + (1 - a_n)\rho_n', \tag{2.79}$$

where

$$\rho_n' = \begin{cases} \sum_{i=n+1}^{\infty} p_i/(1-a_n)|i\rangle\langle i| & (a_n < 1) \\ \rho_n & (a_n = 1) \end{cases}.$$

Then the linearity of $\mathcal{T}$ shows

$$\langle\Phi|\mathcal{T}(\rho_0)|\Psi\rangle = a_n\langle\Phi|\mathcal{T}(\rho_n)|\Psi\rangle + (1-a_n)\langle\Phi|\mathcal{T}(\rho_n')|\Psi\rangle.$$

Thus, if $\langle\Phi|\mathcal{T}(\rho_n')|\Psi\rangle$ is finite in $n \to \infty$ limit, then the lemma is established since

$$\lim_{n\to\infty} a_n = 1.$$

For the purpose of proving the finiteness of $\langle\Phi|\mathcal{T}(\rho_n')|\Psi\rangle$, it is sufficient to show that $|\langle\Phi|\tilde{\rho}|\Psi\rangle|$ is bounded from above by $\|\Phi\|\ \|\Psi\|$ for an arbitrary element $\tilde{\rho}$ of $\mathcal{P}(\tilde{\mathcal{F}})$. This is easy to prove as follows.

$$|\langle\Phi|\tilde{\rho}|\Psi\rangle| = |\sum_i \tilde{p}_i\langle\Phi|\tilde{i}\rangle\langle\tilde{i}|\Psi\rangle| \le \sum_i |\langle\Phi|\tilde{i}\rangle\langle\tilde{i}|\Psi\rangle| \le \|\Phi\|\ \|\Psi\|, \tag{2.80}$$

where we have used a diagonal decomposition

$$\tilde{\rho} = \sum_i \tilde{p}_i |\tilde{i}\rangle\langle\tilde{i}|.$$

□

**Theorem 7** *Assume the following three assumptions: **a**. $\mathcal{T}$ is a linear map of $\mathcal{P}(\mathcal{F})$ into $\mathcal{P}(\tilde{\mathcal{F}})$, **b**. $f$ is a continuous function convex to below and there are non-negative constants $c_1$, $c_2$ and $c_3$ such that $|f((1-\epsilon)x) - f(x)| \le |\epsilon|(c_1|f(x)| + c_2|x| + c_3)$ for $^{\forall}x$ $(\ge 0)$ and sufficiently small $|\epsilon|$, **c**. there are positive definite density matrices $\rho_\infty$ $(\in \mathcal{P}(\mathcal{F}))$ and $\tilde{\rho}_\infty$ $(\in \mathcal{P}(\tilde{\mathcal{F}}))$ such that $\mathcal{T}(\rho_\infty) = \tilde{\rho}_\infty$.*

*If* $[\rho_\infty, \rho_0] = [\tilde{\rho}_\infty, \mathcal{T}(\rho_0)] = 0$ *and* $\boldsymbol{Tr}[\rho_\infty | f(\rho_0 \rho_\infty^{-1})|] < \infty$, *then*

$$\tilde{\mathcal{U}}\left[\mathcal{T}(\rho_0)\right] \geq \mathcal{U}\left[\rho_0\right], \tag{2.81}$$

*where*

$$\begin{aligned} \mathcal{U}\left[\rho\right] &\equiv -\mathbf{Tr}\left[\rho_\infty f\left(\rho\rho_\infty^{-1}\right)\right], \\ \tilde{\mathcal{U}}\left[\tilde{\rho}\right] &\equiv -\mathbf{Tr}\left[\tilde{\rho}_\infty f\left(\tilde{\rho}\tilde{\rho}_\infty^{-1}\right)\right]. \end{aligned} \tag{2.82}$$

As stated in the first paragraph of this subsection, theorem 7 is used in subsection 6 to prove the generalized second law for a quasi-stationary black hole which arises from gravitational collapse.

**Proof**

First let us decompose the density matrices diagonally as follows:

$$\begin{aligned} \rho_0 &= \sum_{i=1}^{\infty} p_i |i\rangle\langle i|, \quad \rho_\infty = \sum_{i=1}^{\infty} q_i |i\rangle\langle i|, \\ \mathcal{T}(\rho_0) &= \sum_{i=1}^{\infty} \tilde{p}_i |\tilde{i}\rangle\langle\tilde{i}|, \quad \mathcal{T}(\rho_\infty) = \sum_{i=1}^{\infty} \tilde{q}_i |\tilde{i}\rangle\langle\tilde{i}|. \end{aligned} \tag{2.83}$$

Then by lemma 6 and (2.76),

$$\tilde{p}_i = \langle\tilde{i}|\mathcal{T}(\rho_0)|\tilde{i}\rangle = \lim_{n\to\infty} \sum_{j=1}^{n} A_{ij} p_j / a_n, \tag{2.84}$$

where $a_n \equiv \sum_{i=1}^{n} p_i$ and $A_{ij} \equiv \langle\tilde{i}|\mathcal{T}(|j\rangle\langle j|)|\tilde{i}\rangle$. $A_{ij}$ has the following properties:

$$\sum_{i=1}^{\infty} A_{ij} = 1, \quad 0 \leq A_{ij} \leq 1.$$

Similarly it is shown that

$$\tilde{q}_i = \lim_{n\to\infty} \sum_{j=1}^{n} A_{ij} q_j / b_n,$$

where $b_n \equiv \sum_{i=1}^{n} q_i$. By (2.84) and the continuity of $f$, it is shown that

$$f(\tilde{p}_i/\tilde{q}_i) = \lim_{n\to\infty} f\left(\sum_{j=1}^{n} A_{ij} \frac{p_j/a_n}{\tilde{q}_i}\right). \tag{2.85}$$

Next define $C_i^n$ and $\tilde{C}_i^n$ by

$$C_i^n \equiv \sum_{j=1}^{n} A_{ij} q_j / \tilde{q}_i, \quad \tilde{C}_i^n \equiv C_i^n / a_n, \tag{2.86}$$

then the convex property of $f$ means

$$f(\tilde{p}_i/\tilde{q}_i) \leq \lim_{n\to\infty} \sum_{j=1}^{n} \frac{A_{ij} q_j}{C_i^n \tilde{q}_i} f(\tilde{C}_i^n p_j / q_j)$$

since

$$\sum_{j=1}^{n} \frac{A_{ij} q_j}{C_i^n \tilde{q}_i} = 1, \frac{A_{ij} q_j}{C_i^n \tilde{q}_i} \geq 0.$$

Hence

$$-\tilde{\mathcal{U}}[\mathcal{T}(\rho_0)] = \sum_{i=1}^{\infty} \tilde{q}_i f(\tilde{p}_i/\tilde{q}_i) \leq \sum_{i=1}^{\infty} \lim_{n\to\infty} \sum_{j=1}^{n} \frac{A_{ij}q_j}{C_i^n} f(\tilde{C}_i^n p_j/q_j). \tag{2.87}$$

Since $C_i^n$ and $\tilde{C}_i^n$ satisfy

$$\lim_{n\to\infty} C_i^n = \lim_{n\to\infty} \tilde{C}_i^n = 1,$$

it is implied by the assumption about $f$ that

$$\left| f\left(\frac{\tilde{C}_i^n p_j}{q_j}\right) - f\left(\frac{p_j}{q_j}\right)\right| \leq |1-\tilde{C}_i^n|\,(c_1|f(p_j/q_j)| + c_2 p_j/q_j + c_3)$$

for sufficiently large $n$. Therefore

$$\begin{aligned}
&\left|\sum_{j=1}^{n} \frac{A_{ij}q_j}{C_i^n}\left(f(\tilde{C}_i^n p_j/q_j) - f(p_j/q_j)\right)\right| \\
&\leq \frac{|1-\tilde{C}_i^n|}{C_i^n}\left(c_1\sum_{j=1}^{n} A_{ij}q_j|f(p_j/q_j)| + c_2\sum_{j=1}^{n} A_{ij}p_j + c_3\sum_{j=1}^{n} A_{ij}q_j\right) \\
&\leq \frac{|1-\tilde{C}_i^n|}{C_i^n}\left(c_1\sum_{j=1}^{n} q_j|f(p_j/q_j)| + c_2\sum_{j=1}^{n} p_j + c_3\sum_{j=1}^{n} q_j\right),
\end{aligned}$$

where we have used $0 \leq A_{ij} \leq 1$ to obtain the last inequality. Since the first term in the brace in the last expression is finite in $n \to \infty$ limit by the assumption of the absolute convergence of $\mathcal{U}[\rho_0]$ and all the other terms in the brace are finite,

$$\lim_{n\to\infty}\left|\sum_{j=1}^{n} \frac{A_{ij}q_j}{C_i^n}\left(f(\tilde{C}_i^n p_j/q_j) - f(p_j/q_j)\right)\right| = 0.$$

Moreover, by the absolute convergence of $\mathcal{U}[\rho_0]$, it is easily shown that

$$\lim_{n\to\infty}\left|\left(\frac{1}{C_i^n} - 1\right)\sum_{j=1}^{n} A_{ij}q_j f(p_j/q_j)\right| = 0.$$

Thus

$$-\tilde{\mathcal{U}}[\mathcal{T}(\rho_0)] \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}q_j f(p_j/q_j). \tag{2.88}$$

We can interchange sum over $i$ and sum over $j$ in the right hand side of (2.88) since it converges absolutely by the absolute convergence of $\mathcal{U}[\rho_0]$. Hence

$$-\tilde{\mathcal{U}}[\mathcal{T}(\rho_0)] \leq \sum_{j=1}^{\infty} q_j f(p_j/q_j) = -\mathcal{U}[\rho_0].$$

□

**Proof of the generalized second law**

Let us combine theorem 3 with theorem 7 to prove the generalized second law. In theorem 7 set the linear map $\mathcal{T}$, the convex function $f(x)$ and the density matrices $\rho_\infty$ and $\tilde{\rho}_\infty$ as follows.

$$\begin{aligned} \mathcal{T} &= T, \\ f(x) &= x \ln x, \\ \rho_\infty &= \rho_{th}(\beta_{BH}, \Omega_{BH}), \\ \tilde{\rho}_\infty &= \tilde{\rho}_{th}(\beta_{BH}, \Omega_{BH}). \end{aligned} \tag{2.89}$$

Note that it is theorem 3 that makes such a setting possible. Hence, if an initial state $\rho_0$ and the corresponding final state $T(\rho_0)$ satisfy

$$[\rho_0, \rho_{th}(\beta_{BH}, \Omega_{BH})] = [T(\rho_0), \tilde{\rho}_{th}(\beta_{BH}, \Omega_{BH})] = 0 \tag{2.90}$$

and $\mathcal{U}[\rho_0]$ converges absolutely, theorem 7 can be applied to the system of the quasi-stationary black hole and the scalar field around it. Now

$$\begin{aligned} \mathcal{U}[\rho_0] &= -\boldsymbol{Tr}\left[\rho_0 \ln \rho_0\right] - \beta_{BH}\left(\boldsymbol{Tr}\left[\boldsymbol{E}\rho_0\right] - \Omega_{BH}\boldsymbol{Tr}\left[\boldsymbol{L}_z\rho_0\right]\right) - \ln Z \\ &= U\left[\rho_0; \beta_{BH}, \Omega_{BH}\right) - \ln Z, \\ \tilde{\mathcal{U}}[\mathcal{T}(\tilde{\rho}_0)] &= -\boldsymbol{Tr}\left[\tilde{\rho}_0 \ln \tilde{\rho}_0\right] - \beta_{BH}\left(\boldsymbol{Tr}\left[\tilde{\boldsymbol{E}}\tilde{\rho}_0\right] - \Omega_{BH}\boldsymbol{Tr}\left[\tilde{\boldsymbol{L}}_z\tilde{\rho}_0\right]\right) - \ln Z \\ &= U\left[T(\tilde{\rho}_0); \beta_{BH}, \Omega_{BH}\right) - \ln Z, \end{aligned} \tag{2.91}$$

where

$$\begin{aligned} \boldsymbol{E} &\equiv \sum_{\{n_{i\gamma}\}} \left(\sum_i n_{i\gamma}\omega_i\right) |\{n_{i\gamma}\}\rangle\langle\{n_{i\gamma}\}|, \\ \boldsymbol{L}_z &\equiv \sum_{\{n_{i\gamma}\}} \left(\sum_i n_{i\gamma}m_i\right) |\{n_{i\gamma}\}\rangle\langle\{n_{i\gamma}\}|, \end{aligned}$$

and

$$\begin{aligned} \tilde{\boldsymbol{E}} &\equiv \sum_{\{n_{i\rho}\}} \left(\sum_i n_{i\rho}\omega_i\right) |\{n_{i\rho}\}\rangle\langle\{n_{i\rho}\}|, \\ \tilde{\boldsymbol{L}}_z &\equiv \sum_{\{n_{i\rho}\}} \left(\sum_i n_{i\rho}m_i\right) |\{n_{i\rho}\}\rangle\langle\{n_{i\rho}\}|. \end{aligned}$$

Thus the inequality (2.81) in this case is (2.74) itself, which in turn is equivalent to the generalized second law. Finally, theorem 7 proves the generalized second law for a quasi-stationary black hole which is formed by gravitational collapse, provided that an initial density matrix $\rho_0$ of the scalar field satisfies the above assumptions. For example, it is guaranteed by lemma 2 that if $\rho_0$ is diagonal in the basis $\{|\{n_{i\gamma}\}\rangle\}$ then $T(\rho_0)$ is also diagonal in the basis $\{|\{n_{i\rho}\}\rangle\}$ and (2.90) is satisfied. The assumption of the absolute convergence of $U[\rho_0; \beta_{BH}, \Omega_{BH})$ holds whenever initial state $\rho_0$ at $\mathcal{I}^-$ contains at most finite number of excitations. Therefore the assumptions are satisfied when $\rho_0$ is diagonal in the basis $\{|\{n_{i\gamma}\}\rangle\}$ and contains at most finite number of excitations.

Here we have to admit that $\ln Z$ in Eqs. (2.91) diverges formally due to the infinite volume of the system. However, it should be possible to avoid this divergence by a suitable regularization. It will be valuable to analyze what kind of regularization does well.

### 2.3.3 Concluding remark

Now we make a comments on Frolov and Page's statement that their proof of the generalized second law may be applied to the case of the black hole formed by gravitational collapse [52]. Their proof for a quasi-stationary eternal black hole is based on the following two assumptions: (1) a state of matter fields on the past horizon is thermal one; (2) a set of radiation modes on the past horizon and a set of modes on the past null infinity are quantum mechanically uncorrelated. These two assumptions are reasonable for the eternal case since a black hole emit a thermal radiation. In the case of a black hole formed by gravitational collapse, we might expect that things would go well by simply replacing the past horizon with a null surface at a moment of a formation of a horizon ($v = v_0$ surface in *Figure 2.1*). However, a state of the matter fields on the past horizon is completely determined by a state of the fields before the horizon formation ($v < v_0$ in *Figure* 2.1), in which there is no causal effect of the existence of future horizon. Since the essential origin of the thermal radiation from a black hole is the existence of the horizon, the state of the fields on the null surface has not to be a thermal one. Hence the assumption (1) does not hold in this case. Although the above replacement may be the most extreme one, a replacement of the past horizon by an intermediate null hypersurface causes an intermediate violation of the assumption (1) and (2) due to the correlation between modes on the future null infinity and modes on the past null infinity located after the horizon formation. The correlation can be seen in (2.65) explicitly in the case of the replacement by the future null infinity. Thus we conclude that their proof can not be applied to the case of the black hole formed by a gravitational collapse.

Finally we discuss a generalization of our proof to a dynamical background. For the case of a dynamical background, $\beta_{BH}$ and $\Omega_{BH}$ are changed from time to time by a possible backreaction. Thus, to prove the generalized second law for the dynamical background, we have to generalize theorem 3 to the dynamical case consistently with the backreaction. Once this can be achieved, theorem 7, combined with the quasi-local first law of black hole dynamics derived in subsection 2.2.2, seems useful to prove the generalized second law for the dynamical background.

Here we have to admit that, in the dynamical situation, entropy $S_{matter}$ of matter fields might lose its physical meaning, while dynamical black hole entropy can be defined as a quarter of trapping horizon area (see subsection 2.2.1). However, if the dynamical version of the generalized second law would be proved for some definition of $S_{matter}$ then, by integrating it from an initial stationary state to a final stationary state, we would be able to obtain the generalized second law for finite changes of black hole parameters. (Note that in the proof given in this section, it has been assumed implicitly by the quasi-stationarity that $\Delta S_{BH} \ll S_{BH}$, $\Delta M_{BH} \ll M_{BH}$, etc.) For example, if we take a black hole as the initial state and a flat spacetime as the final one, then the finite version of the generalized second law insists that matter entropy is increased by evaporation of the initial black hole and that the produced entropy is greater than the initial black hole entropy. Thus, the generalization is necessary for a detailed investigation of the information loss problem.

# Chapter 3

# Black hole entropy

## 3.1 D-brane statistical-mechanics

### 3.1.1 Black brane solution in the type IIB superstring

The low-energy effective theory of type IIB superstring contains, as its bosonic part, a metric field, a dilaton field, a R-R field and NS-NS fields. By setting the NS-NS fields to be zero, we obtain the following low-energy effective action in the 10-dimensional Einstein frame.

$$\frac{1}{16\pi G_{10}}\int d^{10}x\sqrt{-g}[R-\frac{1}{2}(\nabla\phi)^2-\frac{1}{12}e^{\phi}H^2], \tag{3.1}$$

where $\phi$ and $H$ are the dilaton field and the R-R three-form field strength, respectively. (The 10-dimensional Newton's constant $G_{10}$ is defined so that the dilaton $\phi$ vanishes asymptotically.)

For this effective theory we consider toroidal compactification to five dimensions with an $S^1$ of length $2\pi R$, a $T^4$ of four-volume $(2\pi)^4V$, and momentum along the $S^1$: we assume the 10-dimensional metric of the form

$$ds^2_{(10)}=e^{-2(4\chi+\psi)/3}g^{(5)}_{\mu\nu}dx^\mu dx^\nu+e^{2\psi}(dX_5+A_\mu dx^\mu)^2+e^{2\chi}\sum_{i=6}^{9}(dX_i)^2, \tag{3.2}$$

where $\mu=0,1,\cdots,4$, $i=6,\cdots,9$, and all fields depend only on $x^\mu$. Here we also assume that $X_5$ is periodically identified with period $2\pi R$ and that each of $X_i$ is identified with period $2\pi V^{1/4}$. Note that the conformal factor $e^{-2(4\chi+\psi)/3}$ makes the metric $g^{(5)}_{\mu\nu}dx^\mu dx^\nu$ be in the 5-dimensional Einstein frame, and that the field $A_\mu$ becomes a $U(1)$ gauge field in 5-dimensions.

In Ref. [68] a six parameter family of black brane solutions of the equation of motion following from the action (3.1) was given. The metric is of the form (3.2) with

$$\begin{aligned}
g^{(5)}_{\mu\nu}dx^\mu dx^\nu &= -f^{-2/3}\left(1-\frac{r_0^2}{r^2}\right)dt^2+f^{1/3}\left[\left(1-\frac{r_0^2}{r^2}\right)^{-1}dr^2+r^2d\Omega^2_{(3)}\right],\\
A_\mu dx^\mu &= f_3^{-1}\frac{r_0^2\sinh 2\sigma}{2r^2}dt,\\
e^{2\chi} &= f_1^{1/4}f_2^{-1/4},\\
e^{2\psi} &= f_1^{-3/4}f_2^{-1/4}f_3,
\end{aligned} \tag{3.3}$$

where

$$\begin{aligned} f_1 &= 1 + \frac{r_0^2 \sinh^2 \alpha}{r^2}, \\ f_2 &= 1 + \frac{r_0^2 \sinh^2 \gamma}{r^2}, \\ f_3 &= 1 + \frac{r_0^2 \sinh^2 \sigma}{r^2}, \\ f &= f_1 f_2 f_3. \end{aligned} \tag{3.4}$$

The solution is parameterized by the six independent quantities $\alpha$, $\gamma$, $\sigma$, $r_0$, $R$ and $V$.

As seen as a black hole in 5-dimensions, this black brane solution has two charges $Q_1$ and $Q_2$ associated with the R-R field $H$ and the charge $N$ associated with the Kaluza-Klein gauge field $A_\mu$. These charges are written in terms of the six parameters as

$$\begin{aligned} Q_1 &= \frac{V r_0^2}{2g} \sinh 2\alpha, \\ Q_5 &= \frac{r_0^2}{2g} \sinh 2\gamma, \\ N &= \frac{R^2 V r_0^2}{2g^2} \sinh 2\sigma, \end{aligned} \tag{3.5}$$

where $g$ is the 10-dimensional string coupling, by which the 10-dimensional Newton's constant is written as $G_{10} = 8\pi^6 g^2$. Note that, from the general argument of Kaluza-Klein compactification, the charge $N$ is related to the momentum $P$ around the $S^1$ as $P = N/R$. The 5-dimensional black hole also has charges $P_\psi$ and $P_\chi$ related to the asymptotic fall-off of $\psi$ and $\chi$, which are given by

$$\begin{aligned} P_\psi &= \frac{RV r_0^2}{2g^2} [\cosh 2\sigma - \frac{1}{2}(\cosh 2\alpha + \cosh 2\gamma)], \\ P_\chi &= \frac{RV r_0^2}{2g^2} (\cosh 2\alpha - \cosh 2\gamma). \end{aligned} \tag{3.6}$$

In 10-dimensions $P_\psi$ and $P_\chi$ are pressures which describe how the energy changes for isentropic variations in $R$ and $V$. The ADM energy associated with this 5-dimensional black hole is

$$E = \frac{RV r_0^2}{2g^2} (\cosh 2\alpha + \cosh 2\gamma + \cosh 2\sigma). \tag{3.7}$$

Hence the black brane solution can be parameterized by the five charges ($Q_1$, $Q_5$, $N$, $P_\psi$, $P_\chi$) and the ADM energy $E$.

For this solution as a 5-dimensional black hole the Bekenstein-Hawking entropy $S_{BH}$ and the Hawking temperature $T_{BH}$ are given by

$$\begin{aligned} S_{BH} &\equiv \frac{A}{4G_5} = 2\pi(\sqrt{n_L} + \sqrt{n_R})(\sqrt{N_1} + \sqrt{N_{\bar{1}}})(\sqrt{N_5} + \sqrt{N_{\bar{5}}}), \\ \frac{1}{T_{BH}} &= \frac{\pi R}{2} \left( \frac{1}{\sqrt{n_L}} + \frac{1}{\sqrt{n_R}} \right) (\sqrt{N_1} + \sqrt{N_{\bar{1}}})(\sqrt{N_5} + \sqrt{N_{\bar{5}}}), \end{aligned} \tag{3.8}$$

where $A$ is the area of the horizon and $G_5$ is the 5-dimensional Newton's constant given by $G_{10} = G_5 (2\pi)^5 RV$. Here $N_1$, $N_{\bar{1}}$, $N_5$, $N_{\bar{5}}$, $n_L$ and $n_R$ are defined by

$$Q_1 = N_1 - N_{\bar{1}},$$

$$
\begin{aligned}
Q_5 &= N_5 - N_{\bar{5}}, \\
N &= n_L - n_R, \\
P_\psi &= -\frac{R}{2g}(N_1 + N_{\bar{1}}) - \frac{RV}{2g}(N_5 + N_{\bar{5}}) + \frac{1}{R}(n_L + n_R), \\
P_\chi &= \frac{R}{g}(N_1 + N_{\bar{1}}) - \frac{RV}{g}(N_5 + N_{\bar{5}}), \\
E &= \frac{R}{g}(N_1 + N_{\bar{1}}) + \frac{RV}{g}(N_5 + N_{\bar{5}}) + \frac{1}{R}(n_L + n_R).
\end{aligned} \tag{3.9}
$$

The scales of compactification are written in terms of these quantities as

$$
\begin{aligned}
R &= \left(\frac{g^2 n_L n_R}{N_1 N_{\bar{1}}}\right)^{1/4}, \\
V &= \left(\frac{N_1 N_{\bar{1}}}{N_5 N_{\bar{5}}}\right)^{1/2}.
\end{aligned} \tag{3.10}
$$

The black brane solution characterized by ($N_1$, $N_{\bar{1}}$, $N_5$, $N_{\bar{5}}$, $n_L$, $n_R$) can be interpreted as a configuration of strings and solitons in the type IIB superstring theory [68], provided that interactions among string and soliton can be neglected. The configuration is composed of $N_1$ "constituent" D-strings wrapped on the $S^1$, $N_{\bar{1}}$ "constituent" anti-D-strings wrapped on the $S^1$, $N_5$ "constituent" D-fivebranes wrapped on the $T^5 = T^4 \times S^1$, and $N_{\bar{5}}$ "constituent" anti-D-fivebranes wrapped on the $T^5$. Open strings on the D-branes have momentum along the $S^1$ so that $P_L \equiv n_L/R$ is the momentum along the $S^1$ which is a sum over left-moving massless modes of open strings and that $P_R \equiv n_R/R$ is the momentum which is a sum over right-moving massless modes. This interpretation is based on the following three facts:

(i) A single-wound D-string or anti-D-string has

$$
\begin{aligned}
Q_1 = \pm 1, Q_5 = 0, N = 0, \\
P_\psi = -\frac{R}{2g}, P_\chi = \frac{R}{g}, E = \frac{R}{g},
\end{aligned} \tag{3.11}
$$

where plus sign is for the D-string and minus sign is for the anti-D-string.

(ii) A single-wound D-fivebrane or anti-D-fivebrane has

$$
\begin{aligned}
Q_1 = 0, Q_5 = \pm 1, N = 0, \\
P_\psi = -\frac{RV}{2g}, P_\chi = -\frac{RV}{g}, E = \frac{RV}{g},
\end{aligned} \tag{3.12}
$$

where plus sign is for the D-fivebrane and minus sign is for the anti-D-fivebrane.

(iii) A left- or right- moving string with momentum $P = \pm n/R$ along the $S^1$ has

$$
\begin{aligned}
Q_1 = 0, Q_5 = 0, N = \pm n, \\
P_\psi = \frac{n}{R}, P_\chi = 0, E = \frac{n}{R},
\end{aligned} \tag{3.13}
$$

where plus sign is for the left-mover and minus sign is for the right-mover.

In the following arguments, we consider the case that there are no anti-D-branes: $N_{\bar{1}} = N_{\bar{5}} = 0$, and thus $Q_1 = N_1$, $Q_5 = N_5$. In this case the Bekenstein-Hawking

entropy and the Hawking temperature are given by

$$\begin{aligned} S_{BH} &= 2\pi(\sqrt{n_L}+\sqrt{n_R})\sqrt{Q_1Q_5}, \\ \frac{1}{T_{BH}} &= \frac{\pi R}{2}\left(\frac{1}{\sqrt{n_L}}+\frac{1}{\sqrt{n_R}}\right)\sqrt{Q_1Q_5}. \end{aligned} \tag{3.14}$$

The purpose of the remaining part of this section is to explain these expressions of the entropy and the temperature by using the D-brane technology.

### 3.1.2 Number of microscopic states

Consider a set of $Q_1$ single-wound D-strings wrapped on the $S^1$ and a set of $Q_5$ single-wound D-fivebranes wrapped on $T^5 = T^4 \times S^1$. The D-strings may be connected up to form a set of multiply-wound D-strings, which is composed of $N_{q_1}^{(1)}$ D-strings of length $2\pi R q_1$ ($q_1 = 1, 2, \cdots$). The D-fivebranes may also be connected up to form a set of multiply-wound D-fivebranes, which is composed of $N_{q_5}^{(5)}$ D-fivebranes of length $2\pi R q_5$ ($q_5 = 1, 2, \cdots$) along the $S^1$. This general configuration of D-strings and D-fivebranes was first considered in Ref. [43]. Note that, by the conservation of charges $Q_1$ and $Q_5$, the integers $N_{q_1}^{(1)}$ and $N_{q_5}^{(5)}$ must satisfy the following constraints:

$$\begin{aligned} Q_1 &= \sum_{q_1} q_1 N_{q_1}^{(1)}, \\ Q_5 &= \sum_{q_5} q_5 N_{q_5}^{(5)}, \end{aligned} \tag{3.15}$$

since a D-string which winds $q_1$ times has charge $(Q_1, Q_5)=(q_1, 0)$ and a D-fivebrane which winds $q_5$ times has charge $(Q_1, Q_5)=(0, q_5)$.

For example, the configuration considered by Callan and Maldacena [70] is the case that $N_1^{(1)} = Q_1$, $N_1^{(5)} = Q_5$ and all other $N_{q_1}^{(1)}$ and $N_{q_5}^{(5)}$ are zero: no D-branes are connected up. (*Figure* 3.1(a) is a schematic picture of this situation for $Q_1 = 2$, $Q_5 = 3$.) On the contrary Maldacena and Susskind [69] considered the configuration such that $N_{Q_1}^{(1)} = N_{Q_5}^{(5)} = 1$ and all other $N_{q_1}^{(1)}$ and $N_{q_5}^{(5)}$ are zero: one long D-string and one long D-fivebrane. (*Figure* 3.1(b) is a schematic picture of this situation for $Q_1 = 2$, $Q_5 = 3$.) The general class of configurations introduced above includes more abundant situations. For example, the situation in *Figure* 3.1(c) is an example for $Q_1 = 2$, $Q_5 = 3$: $N_2^{(1)} = N_1^{(5)} = N_2^{(5)} = 1$ and others are zero.

As explained in the previous subsection, the charges $n_L$ and $n_R$ are implemented in the D-brane picture so that $P_L \equiv n_R/R$ is the momentum along the $S^1$ which is a sum over left-moving massless modes of open strings on the D-branes and that $P_R \equiv n_L/R$ is the momentum which is a sum over right-moving massless modes.

There are several types of open strings on the D-brane configuration: those with both boundaries being on a common D-string; those with one boundary being on a D-string and another boundary being on a different D-string; those with one boundary being on a D-string and another boundary being on a D-fivebrane; etc. Among them we only consider a set of strings each of which connects a D-string and a D-fivebrane, since if many such strings are excited then all other strings become massive [71]. As in the situation considered in Ref. [70], each string connecting a D-string and a D-fivebrane has 4 bosonic and 4 fermionic degrees of freedom. Hence we have $4N_{q_1}^{(1)}N_{q_5}^{(5)}$ bosonic and $4N_{q_1}^{(1)}N_{q_5}^{(5)}$ fermionic degrees of freedom for strings with one boundary on one of $N_{q_1}^{(1)}$ $q_1$-wound D-strings and another boundary on one of $N_{q_5}^{(5)}$ $q_5$-wound D-fivebranes.

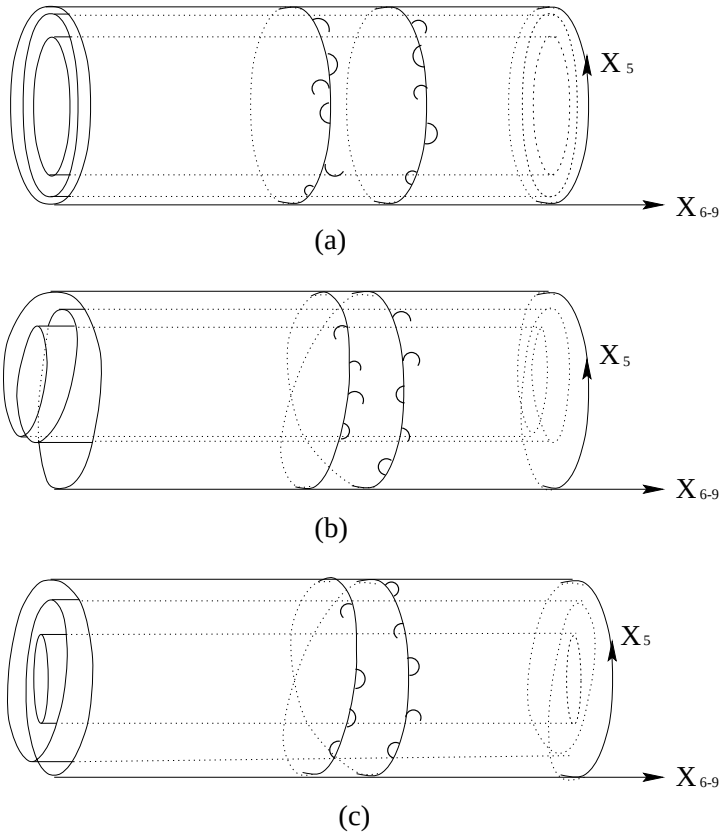


Figure 3.1: (a) The situation considered by Callan and Maldacena is shown for $Q_1 = 2, Q_5 = 3$. There are three D-fivebranes wrapped on $T^5$ ($X_{5-9}$ are coordinates for the torus.) and two D-strings wrapped on $S^1$ of the torus. ($X_5$ is a coordinate for the $S^1$.) (b) The situation considered by Maldacena and Susskind is shown for $Q_1 = 2$, $Q_5 = 3$. There are a triple-wound D-fivebrane and a double-wound D-string. (c) Another example of the generalized configurations of D-branes for $Q_1 = 2$, $Q_5 = 3$.

For each of fermionic or bosonic degrees of freedom of the string connecting a $q_1$-wound D-string and a $q_5$-wound D-fivebrane, the spectrum of the momentum along the $S^1$ is

$$p_n(q_1,q_5) = \frac{\pm n}{R(q_1,q_5)_{LCM}}, \quad (n=1,2,\cdots) \tag{3.16}$$

$$('+' \text{ for left moving} \quad ; \quad '-' \text{ for right moving}), \tag{3.17}$$

where $(q_1,q_5)_{LCM}$ is the least common multiple of $q_1$ and $q_5$, since the boundary condition of the strings is

$$X^5 \sim X^5 + 2\pi R(q_1,q_5)_{LCM}. \tag{3.18}$$

We regard the momentum $P_L \equiv n_L/R$ ($P_R \equiv -n_R/R$) along the $S^1$ as a sum over all left (right) moving massless modes of the strings. It is evident from the assumption of neglecting interactions between strings that the number $d(n_L,n_R)$ of string states consistent with given values of $n_L$ and $n_R$ is a product of the numbers $d(n_L)$ and $d(n_R)$ of left- and right-moving string states:

$$d(n_L,n_R) = d(n_L)d(n_R). \tag{3.19}$$

At this point it is easy to show that $d(n_L)$ satisfies the following relation for an arbitrary value of $w$ satisfying $0<w<1$.

$$\sum_{l=0}^{\infty} d(n_L) w^{n_L} = \prod_{q_1}\prod_{q_5}\prod_{n=1}^{\infty}\left[\frac{1+w^{n/(q_1,q_5)_{LCM}}}{1-w^{n/(q_1,q_5)_{LCM}}}\right]^{4N_{q_1}^{(1)}N_{q_5}^{(5)}}, \tag{3.20}$$

where $n_L = l/Q_1!Q_5!$. It is evident that $d(0)=1$. Note that the denominator in the r.h.s. represents a partition function for the bosonic modes of open strings and the numerator represents a partition function for the fermionic modes.

Note that $d(n_L)$ can be considered as a residue of the r.h.s. of Eq. (3.20) divided by $x^{l+1}$, where $x = w^{1/Q_1!Q_5!}$. Hence, $d(n_L)$ can be written as a complex integral:

$$\begin{aligned} d(n_L) &= \frac{1}{2\pi i}\oint dx x^{-(l+1)} \prod_{q_1}\prod_{q_5}\prod_{n=1}^{\infty}\left[\frac{1+w^{n/(q_1,q_5)_{LCM}}}{1-w^{n/(q_1,q_5)_{LCM}}}\right]^{4N_{q_1}^{(1)}N_{q_5}^{(5)}} \\ &= \frac{1}{2\pi i}\oint dw \exp[h(w)], \end{aligned} \tag{3.21}$$

where

$$h(w) = 4\sum_{q_1,q_5}\sum_{n} N_{q_1}^{(1)}N_{q_5}^{(5)} \ln\left[\frac{1+w^{n/(q_1,q_5)_{LCM}}}{1-w^{n/(q_1,q_5)_{LCM}}}\right] - (n_L+1)\ln w. \tag{3.22}$$

The contour in the integration w.r.t. $x$ is a closed curve surrounding $x=0$ in complex $x$-plane and, the contour in the integration w.r.t. $w$ is a closed curve surrounding $w=0$ in complex $w$-plane.

Now we give an asymptotic formula for $d(n_L)$ in the limit

$$\sqrt{\frac{Q_1Q_5}{n_L}} \ll \min_{(q_1,q_5)\in I}(q_1,q_5)_{LCM}, \tag{3.23}$$

where $I \equiv \{(q_1,q_5)|N_{q_1}^{(1)}N_{q_5}^{(5)} \neq 0\}$. For this purpose we use an asymptotic expression of $h(w)$. Let us consider a function $g(x)$ defined by

$$g(x) = \prod_{n=1}^{\infty}\left[\frac{1+x^n}{1-x^n}\right], \tag{3.24}$$

where $0 < x < 1$ is assumed. For $x \sim 1$, we can estimate $\ln g(x)$ as

$$\begin{aligned} \ln g(x) &= \sum_{n=1}^{\infty}[\ln(1+x^n) - \ln(1-x^n)] \\ &\approx \frac{1}{-\ln x}\int_0^{\infty} dy[\ln(1+e^{-y}) - \ln(1-e^{-y})] \\ &= \frac{1}{-\ln x}\cdot\frac{\pi^2}{4}, \end{aligned} \tag{3.25}$$

where we have estimated the summation w.r.t. $n$ by the integration w.r.t. $y = -n\ln x$. Hence, we obtain the following asymptotic expression of $h(w)$ for $w^{1/m} \sim 1$, where $m = \min_{(q_1,q_5)\in I}(q_1,q_5)_{LCM}$.

$$h(w) \approx \pi^2 \sum_{q_1,q_5} \frac{N_{q_1}^{(1)} N_{q_5}^{(5)}}{-\ln w^{1/(q_1,q_5)_{LCM}}} - (n_L+1)\ln w. \tag{3.26}$$

We shall estimate the integration w.r.t. $w$ in (3.21) by using this expression and the suddle-point method. When the condition (3.23) is satisfied, it is shown by using this asymptotic expression of $h(w)$ that $exp[h(w)]$ has a saddle point given by

$$-\ln w^{1/m} \approx \frac{\pi}{m}\sqrt{\frac{\sum_{q_1,q_5} N_{q_1}^{(1)} N_{q_5}^{(5)}(q_1,q_5)_{LCM}}{n_L+1}} \qquad \left(\leq \frac{\pi}{m}\sqrt{\frac{Q_1Q_5}{n_L+1}}\right). \tag{3.27}$$

Note that $w^{1/m} \sim 1$ is guaranteed for this saddle point by the condition (3.23). Thus, we obtain the following asymptotic formula of $d(n_L)$ by using the saddle point method.

$$d(n_L) \approx \exp\left[2\pi\sqrt{n_L \sum_{q_1}\sum_{q_5} N_{q_1}^{(1)} N_{q_5}^{(5)}(q_1,q_5)_{LCM}}\right]. \tag{3.28}$$

From this expression it is easily shown that

$$d(n_L) \leq \exp(2\pi\sqrt{n_L Q_1 Q_5}), \tag{3.29}$$

where the bound is saturated if and only if all $(q_1,q_5)$ $(\in I)$ are relatively prime.

Therefore, for a configuration satisfying

$$\sqrt{\frac{Q_1Q_5}{\min(n_L,n_R)}} \ll \min_{(q_1,q_5)\in I}(q_1,q_5)_{LCM}, \tag{3.30}$$

$\ln d(n_L,n_R)$ is bounded from above by the Bekenstein-Hawking entropy given by (3.14):

$$\ln d(n_L,n_R) \leq S_{BH}, \tag{3.31}$$

where the bound is saturated if and only if all $(q_1,q_5)$ $(\in I)$ are relatively prime. This is one of the main results in Ref. [43] and is a generalization of the results in Refs. [70, 69] to the more abundant configurations of D-strings and D-fivebranes.

When $n_R = 0$, which corresponds to an extremal black hole, the condition (3.30) does not hold since the l.h.s diverges. Nonetheless, even in this case, the bound (3.31) does hold if the condition (3.23) is satisfied, since $d(0) = 1$. The bound is saturated if and only if all $(q_1,q_5)$ $(\in I)$ are relatively prime.

### 3.1.3 Canonical ensemble of open strings

The Hawking process can be described in the D-brane picture as a decay process of non-BPS excitations of D-branes [70, 72, 73]. A collision of a right-moving open string with a left-moving one on the D-branes results in an emission of a closed string leaving away from the D-branes, which is interpreted as Hawking radiation. The spectrum of the emission of the closed string can, in principle, be obtained from decay rates of the string, provided that the initial state of the open strings on the D-branes is given. On the contrary, when we do not know anything about the initial state of the open strings, it seems that the best way to evaluate the spectrum is summing up decay rates over all consistent initial states of open strings on the D-branes. Formally, the summation should be done by using a microcanonical ensemble with fixed $n_L$ and $n_R$. Since this ideal summation is not easy, we adopt an approximation. Our strategy of the approximation to perform the summation is to replace the microcanonical ensemble by the canonical ensemble, in which each expectation value of $n_L$ and $n_R$ coincides with the given fixed value in the original microcanonical ensemble. It is easy to see that the corresponding spectrum of closed strings is approximately the thermal one with temperature being the same as the temperature of this canonical ensemble.

Thus, in this subsection we consider a canonical ensemble of open strings on the D-brane configuration introduced in the beginning of subsection 3.1.2. From arguments in the derivation of Eqs. (3.19) and (3.20), we can obtain the partition function of the canonical ensemble as

$$\begin{aligned} Z(\beta,\alpha) &= Z_L(\beta,\alpha)Z_R(\beta,\alpha), \\ Z_{L,R}(\beta,\alpha) &= \prod_{q_1}\prod_{q_5}\prod_{n=1}^{\infty}\left[\frac{1+e^{-\beta e_n(q_1,q_5)-\alpha p_n(q_1,q_5)}}{1-e^{-\beta e_n(q_1,q_5)-\alpha p_n(q_1,q_5)}}\right]^{4N_{q_1}^{(1)}N_{q_5}^{(5)}}, \end{aligned} \tag{3.32}$$

where $\beta$ is inverse temperature and $\alpha$ is chemical potential w.r.t. the total momentum along the $S^1$. Here $p_n$ is given by (3.16) (the plus sign is for $Z_L$ and the minus sign is for $Z_R$) and $e_n = |p_n|$. The quantities $\beta$ and $\alpha$ should be determined by requiring

$$\begin{aligned} \frac{1}{R}(n_L+n_R) &= -\frac{\partial \ln Z(\beta,\alpha)}{\partial\beta}, \\ \frac{1}{R}(n_L-n_R) &= -\frac{\partial \ln Z(\beta,\alpha)}{\partial\alpha}. \end{aligned} \tag{3.33}$$

Note that these requirements originate from the last and the third equalities in (3.9), respectively.

By using the formula (3.25), we can obtain an asymptotic expression of $Z_{L,R}(\beta,\alpha)$ in the limit (3.30). The result is

$$\begin{aligned} \ln Z_{L,R}(\beta,\alpha) &\approx \frac{\pi^2 R}{\beta\pm\alpha}\sum_{q_1}\sum_{q_5} N_{q_1}^{(1)}N_{q_5}^{(5)}(q_1,q_5)_{LCM}, \\ (\text{'}+\text{'} \text{ for } \text{'}L\text{'} \quad &; \quad \text{'}-\text{'} \text{ for } \text{'}R\text{'}). \end{aligned} \tag{3.34}$$

From the condition (3.33), $\beta$ and $\alpha$ are determined as

$$\begin{aligned} \beta+\alpha &\approx \pi R\sqrt{\frac{\sum_{q_1}\sum_{q_5} N_{q_1}^{(1)}N_{q_5}^{(5)}(q_1,q_5)_{LCM}}{n_R}}, \\ \beta-\alpha &\approx \pi R\sqrt{\frac{\sum_{q_1}\sum_{q_5} N_{q_1}^{(1)}N_{q_5}^{(5)}(q_1,q_5)_{LCM}}{n_L}}. \end{aligned} \tag{3.35}$$

Thus, for a configuration satisfying (3.30), the inverse temperature $\beta$ and the entropy $S$ of the effective canonical ensemble are

$$\begin{aligned}
\beta &\approx \frac{\pi R}{2}\left(\frac{1}{\sqrt{n_L}}+\frac{1}{\sqrt{n_R}}\right)\sqrt{\sum_{q_1}\sum_{q_5} N^{(1)}_{q_1}N^{(5)}_{q_5}(q_1,q_5)_{LCM}}, \\
S &= \ln Z(\beta,\alpha) - \beta\frac{\partial \ln Z(\beta,\alpha)}{\partial\beta} - \alpha\frac{\partial \ln Z(\beta,\alpha)}{\partial\alpha} \\
&\approx 2\pi\left(\sqrt{n_L}+\sqrt{n_R}\right)\sqrt{\sum_{q_1}\sum_{q_5} N^{(1)}_{q_1}N^{(5)}_{q_5}(q_1,q_5)_{LCM}} \\
&\approx \ln d(n_L,n_R), \qquad (3.36)
\end{aligned}$$

where $d(n_L,n_R)$ is the number of micro-states evaluated by using the microcanonical ensemble in the previous subsection. In fact, the last equality is trivial from the general arguments in statistical mechanics: entropy in microcanonical ensemble can be calculated approximately by the corresponding canonical ensemble. The temperature $\beta^{-1}$ is bounded from below by the Hawking temperature given by (3.14):

$$\beta^{-1} \geq T_{BH}, \qquad (3.37)$$

where the the bound is saturated if and only if all $(q_1,q_5)$ $(\in I)$ are relatively prime. Note that this condition for the equality is the same as that for the equality in (3.31). Here remember that the temperature $\beta^{-1}$ coincides with the temperature of the thermal spectrum of the closed string emission from the D-branes.

### 3.1.4 Summary and speculations

In this section the D-brane statistical-mechanics has been reviewed by using a configuration of D-strings and D-fivebranes introduced in Ref. [43]. We have considered a set of multiply-wound D-strings, which is composed of $N^{(1)}_{q_1}$ D-strings of length $2\pi R q_1$ $(q_1 = 1, 2, \cdots)$ and a set of multiply-wound D-fivebranes, which is composed of $N^{(5)}_{q_5}$ D-fivebranes of length $2\pi R q_5$ $(q_5 = 1, 2, \cdots)$ along the $S^1$. For configurations satisfying (3.30), the number of microscopic states $d(n_L,n_R)$ of open strings on the D-branes is bounded from above by exponential of the Bekenstein-Hawking entropy $S_{BH}$ of the corresponding black hole, and the temperature $\beta^{-1}$ of a decay of D-brane excitations to closed strings is bounded from below by the Hawking temperature $T_{BH}$ of the corresponding black hole: Eqs.(3.31) and (3.37). Note that $d(n_L,n_R)$ and $\beta^{-1}$ are evaluated microscopically while $S_{BH}$ and $T_{BH}$ are defined in terms of macroscopic quantities (area and surface gravity of the horizon of the corresponding black hole). The bounds (3.31) and (3.37) are saturated if and only if all $(q_1,q_5)$ $(\in I)$ are relatively prime.

Thus several speculations may be possible.

- Inside a black hole characterized by the four parameters ($Q_1$, $Q_5$, $n_L$, $n_R$), some dynamical processes may occur. The processes may be described in the D-brane picture: D-branes repeat fission and fusion to settle down to one of the states for which all $(q_1,q_5)$ $(\in I)$ are relatively prime.

- During the process the microscopic entropy increases to reach the Bekenstein-Hawking entropy of the corresponding black hole. Moreover the temperature of closed string radiation from the D-branes decreases to reach the Hawking temperature of the black hole.

These speculations may be significant to investigate the microstates of dynamical black holes. For example let us consider a merger of two black holes $B_1$ and $B_2$

and suppose that $B_1$ corresponds to a D-brane configuration $\{N_{q_1}^{(11)}, N_{q_5}^{(15)}\}$ and $B_2$ corresponds to a configuration $\{N_{q_1}^{(21)}, N_{q_5}^{(25)}\}$. Just after merging, the large black hole $B$ formed by the merger corresponds to a configuration $\{N_{q_1}^{(1)} = N_{q_1}^{(11)} + N_{q_1}^{(21)}, N_{q_5}^{(5)} = N_{q_5}^{(15)} + N_{q_5}^{(25)}\}$, provided that directions of the D-strings are the same for $B_1$ and $B_2$. In general the last configuration does not saturate the bounds (3.31) and (3.37) even if the configurations for $B_1$ and $B_2$ saturate the bounds. Thus, in general just after the merger the microscopic entropy of $B$ does not agree with the corresponding Bekenstein-Hawking entropy and the temperature of the closed string radiation does not agree with the corresponding Hawking temperature. However, after a sufficiently long time the D-branes' fission and fusion settle the entropy and the temperature to the Bekenstein-Hawking entropy and the Hawking temperature.

## 3.2 Brick wall model

The entanglement interpretation, which will be investigated in detail in section 3.3, seems to be implicit in, and is certainly closely related to a pioneering calculation done by Gerard 'tHooft [35] in 1985. He considered the statistical thermodynamics of quantum fields in the Hartle-Hawking state (i.e. having the Hawking temperature $T_{BH}$ at large radii) propagating on a fixed Schwarzschild background of mass $M$. To control divergences, 'tHooft introduced a "brick wall"—actually a static spherical mirror at which the fields are required to satisfy Dirichlet or Neumann boundary conditions—with radius a little larger than the gravitational radius $2M$. He found, in addition to the expected volume-dependent thermodynamical quantities describing hot fields in a nearly flat space, additional wall contributions proportional to the area. These contributions are, however, also proportional to $\alpha^{-2}$, where $\alpha$ is the proper altitude of the wall above the gravitational radius, and thus diverge in the limit $\alpha \to 0$. For a specific choice of $\alpha$ (which depends on the number of fields, etc., but is generally of order $l_{pl}$), 'tHooft was able to recover the Bekenstein-Hawking formula with the correct coefficient.

However, this calculation raises a number of questions which have caused many, including 'tHooft himself, to have reservations about its validity and consistency.

(a) $S_{BH}$ is here obtained as a one-loop effect, originating from thermal excitations of the quantum fields. Does this material contribution to $S_{BH}$ have to be *added* to the zero-loop Gibbons-Hawking contribution which arises from the gravitational part of the action and already by itself accounts for the full value of $S_{BH}$? [74]

(b) The ambient quantum fields were assumed to be in the Hartle-Hawking state. Their stress-energy should therefore be bounded (of order $M^{-4}$ in Planck units) near the gravitational radius, and negligibly small for large masses. However, 'tHooft's calculation assigns to them enormous (Planck-level) energy densities near the wall.

(c) The integrated field energy gives a wall contribution to the mass

$$\Delta M = \frac{3}{8}M \tag{3.38}$$

when $\alpha$ is adjusted to give the correct value of $S_{BH}$. This suggests a substantial gravitational back-reaction [35] and that the assumption of a fixed geometrical background may be inconsistent [74, 75, 76].

Our main purpose in this section is to point out that these difficulties are only apparent and easy to resolve [44]. The basic remark is that the *brick-wall model*

*strictly interpreted does not represent a black hole.* It represents the exterior of a starlike object with a reflecting surface, compressed to nearly (but not quite) its gravitational radius. The ground state for quantum fields propagating around this star is not the Hartle-Hawking state [77] but the Boulware state [78], corresponding to zero temperature, which has a quite different behavior near the gravitational radius.

In subsection 3.2.1 we summarize essential properties of the Boulware and Hartle-Hawking states that play a role in our arguments. In subsection 3.2.2 we sketch the physical essence of the brick-wall model by using a particle description of quantum fields. A systematic treatment of the model is deferred to subsection 3.2.3, in which the results in subsection 3.2.2 are rigorously derived from the quantum field theory in curved spacetimes. In subsection 3.2.4 we propose a complementary principle between the brick wall model and the Gibbons-Hawking instanton. In Appendix A.3, for completeness, we apply the so-called on-shell method to the brick wall model and show that in the on-shell method we might miss some physical degrees of freedom. Hence, we do not adopt the on-shell method in the main body of this thesis.

### 3.2.1 The Boulware and Hartle-Hawking states

It is useful to begin by summarizing briefly the essential properties of the quantum states that will play a role in our discussion.

In a curved spacetime there is no unique choice of time coordinate. Different choices lead to different definitions of positive-frequency modes and different ground states.

In any static spacetime with static (Killing) time parameter $t$, the Boulware state $|B\rangle$ is the one annuled by the annihilation operators $a_{Kill}$ associated with "Killing modes" (positive-frequency in $t$). In an asymptotically flat space, $|B\rangle$ approaches the Minkowski vacuum at infinity.

In the spacetime of a stationary eternal black hole, the Hartle-Hawking state $|HH\rangle$ is the one annuled by $a_{Krus}$, the annihilation operators associated with "Kruskal modes" (positive-frequency in the Kruskal lightlike coordinates $U$, $V$). This state appears empty of "particles" to free falling observers at the horizon, and its stress-energy is bounded there (not quite zero, because of irremovable vacuum polarization effects).

If, just for illustrative purposes, we consider a $(1+1)$-dimensional spacetime, it is easy to give concrete form to these remarks. We consider a spacetime with metric

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)}, \tag{3.39}$$

and denote by $\kappa(r)$ the redshifted gravitational force, i.e., the upward acceleration $a(r)$ of a stationary test-particle reduced by the redshift factor $f^{1/2}(r)$, so that $\kappa(r) = \frac{1}{2}f'(r)$. A horizon is characterized by $r = r_0$, $f(r_0) = 0$, and its surface gravity defined by $\kappa_0 = \frac{1}{2}f'(r_0)$.

Quantum effects induce an effective quantum stress-energy $T_{ab}$ $(a, b, \cdots = r, t)$ in the background geometry (3.39). If we assume no net energy flux ($T^r_t = 0$)—thus excluding the Unruh state—$T_{ab}$ is completely specified by a quantum energy density $\rho = -T^t_t$ and pressure $P = T^r_r$. These are completely determined (up to a boundary condition) by the conservation law $T^b_{a;b} = 0$ and the trace anomaly, which is

$$T^a_a = \frac{\hbar}{24\pi}R \tag{3.40}$$

for a massless scalar field, with $R = -f''(r)$ for the metric (3.39). Integration gives

$$f(r)P(r) = -\frac{\hbar}{24\pi}(\kappa^2(r) + const.). \tag{3.41}$$

Different choices of the constant of integration correspond to different boundary conditions, i.e., to different quantum states.

For the Hartle-Hawking state, we require $P$ and $\rho$ to be bounded at the horizon $r = r_0$, giving

$$\begin{aligned} P_{HH} &= \frac{\hbar}{24\pi}\frac{\kappa_0^2 - \kappa^2(r)}{f(r)}, \\ \rho_{HH} &= P_{HH} + \frac{\hbar}{24\pi}f''(r). \end{aligned} \tag{3.42}$$

When $r \to \infty$ this reduces to (setting $f(r) \to 1$)

$$\begin{aligned} \rho_{HH} &\simeq P_{HH} = \frac{\pi}{6\hbar}T_{BH}^2, \\ T_{BH} &= \hbar\kappa_0/2\pi, \end{aligned} \tag{3.43}$$

which is appropriate for one-dimensional scalar radiation at the Hawking temperature $T_{BH}$.

For the Boulware state, the boundary condition is $P = \rho = 0$ when $r = \infty$. The integration constant in (3.41) must vanish, and we find

$$\begin{aligned} P_B &= -\frac{\hbar}{24\pi}\frac{\kappa^2(r)}{f(r)}, \\ \rho_B &= P_B + \frac{\hbar}{24\pi}f''(r). \end{aligned} \tag{3.44}$$

If a horizon were present, $\rho_B$ and $P_B$ would diverge there to $-\infty$.

For the difference of these two stress tensors,

$$\Delta T_a^b = (T_a^b)_{HH} - (T_a^b)_B, \tag{3.45}$$

(3.42) and (3.44) give the exactly thermal form

$$\Delta P = \Delta\rho = \frac{\pi}{6\hbar}T^2(r), \tag{3.46}$$

where $T(r) = T_{BH}/\sqrt{f(r)}$ is the local temperature in the Hartle-Hawking state. We recall that thermal equilibrium in any static gravitational field requires the local temperature $T$ to rise with depth in accordance with Tolman's law [79]

$$T\sqrt{-g_{00}} = const. \tag{3.47}$$

We have found, for this $(1+1)$-dimensional example, that the Hartle-Hawking state is thermally excited above the zero-temperature (Boulware) ground state to a local temperature $T(r)$ which grows without bound near the horizon. Nevertheless, it is the Hartle-Hawking state which best approximates what a gravitational theorist would call a "vacuum" at the horizon.

These remarks remain at least qualitatively valid in $(3+1)$-dimensions, with obvious changes arising from the dimensionality. In particular, the $(3+1)$-dimensional analogue of (3.46) for a massless scalar field,

$$3\Delta P \simeq \Delta\rho \simeq \frac{\pi^2}{30\hbar^3}T^4(r), \tag{3.48}$$

holds to a very good approximation, both far from the black hole and near the horizon. In the intermediate zone there are deviations, but they always remain bounded [80], and will not affect our considerations.

### 3.2.2 A brief sketch of the brick wall model

We shall briefly sketch the physical essence of the brick-wall model. (A systematic treatment is deferred to subsection 3.2.3)

We wish to study the thermodynamics of hot quantum fields confined to the outside of a spherical star with a perfectly reflecting surface whose radius $r_1$ is a little larger than its gravitational radius $r_0$. To keep the total field energy bounded, we suppose the system enclosed in a spherical container of radius $L \gg r_1$.

It will be sufficiently general to assume for the geometry outside the star a spherical background metric of the form

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2. \tag{3.49}$$

This includes as special cases the Schwarzschild, Reissner-Nordström and de Sitter geometries, or any combination of these.

Into this space we introduce a collection of quantum fields, raised to some temperature $T_\infty$ at large distances, and in thermal equilibrium. The local temperature $T(r)$ is then given by Tolman's law (3.47),

$$T(r) = T_\infty f^{-1/2} \tag{3.50}$$

and becomes very large when $r \to r_1 = r_0 + \Delta r$. We shall presently identify $T_\infty$ with the Hawking temperature $T_{BH}$ of the horizon $r = r_0$ of the exterior metric (3.49), continued (illegitimately) into the internal domain $r < r_1$.

Characteristic wavelengths $\lambda$ of this radiation are small compared to other relevant length-scales (curvature, size of container) in the regions of interest to us. Near the star's surface,

$$\lambda \sim \hbar/T = f^{1/2}\hbar/T_\infty \ll r_0. \tag{3.51}$$

Elsewhere in the large container, at large distances from the star,

$$f \simeq 1, \quad \lambda \simeq \hbar/T_\infty \sim r_0 \ll L. \tag{3.52}$$

Therefore, a particle description should be a good approximation to the statistical thermodynamics of the fields (Equivalently, one can arrive at this conclusion by considering the WKB solution to the wave equation, cf. 'tHooft [35] and subsection 3.2.3.)

For particles of rest-mass $m$, energy $E$, 3-momentum $p$ and 3-velocity $v$ as viewed by a local stationary observer, the energy density $\rho$, pressure $P$ and entropy density $s$ are given by the standard expressions

$$\begin{aligned} \rho &= \mathcal{N}\int_0^\infty \frac{E}{e^{\beta E} - \epsilon}\frac{4\pi p^2 dp}{h^3}, \\ P &= \frac{\mathcal{N}}{3}\int_0^\infty \frac{vp}{e^{\beta E} - \epsilon}\frac{4\pi p^2 dp}{h^3}, \\ s &= \beta(\rho + P). \end{aligned} \tag{3.53}$$

Here, as usual,

$$E^2 - p^2 = m^2, \quad v = p/E, \quad \beta = T^{-1}; \tag{3.54}$$

$\epsilon$ is $+1$ for bosons and $-1$ for fermions and the factor $\mathcal{N}$ takes care of helicities and the number of particle species. The total entropy is given by the integral

$$S = \int_{r_1}^{L} s(r) 4\pi r^2 dr/\sqrt{f}, \tag{3.55}$$

where we have taken account of the proper volume element as given by the metric (3.49). The factor $f^{-1/2}$ does not, however, appear in the integral for the gravitational mass of the thermal excitations [81] (It is canceled, roughly speaking, by negative gravitational potential energy):

$$\Delta M_{therm} = \int_{r_1}^{L} \rho(r) 4\pi r^2 dr. \tag{3.56}$$

The integrals (3.55) and (3.56) are dominated by two contributions for large container radius $L$ and for small $\Delta r = r_1 - r_0$:

(a) A volume term, proportional to $\frac{4}{3}\pi L^3$, representing the entropy and mass-energy of a homogeneous quantum gas in a flat space (since $f \simeq 1$ almost everywhere in the container if $L/r_0 \to \infty$) at a uniform temperature $T_\infty$. This is the result that would have been expected, and we do not need to consider it in detail.

(b) Of more interest is the contribution of gas near the inner wall $r = r_1$, which we now proceed to study further. We shall find that it is proportional to the wall area, and diverging like $(\Delta r)^{-1}$ when $\Delta r \to 0$.

Because of the high local temperatures $T$ near the wall for small $\Delta r$, we may insert the ultrarelativistic approximations

$$E \gg m, \quad p \simeq E, \quad v \simeq 1$$

into the integrals (3.53). This gives

$$P \simeq \frac{1}{3}\rho \simeq \frac{\mathcal{N}}{6\pi^2} T^4 \int_0^\infty \frac{x^3 dx}{e^x - \epsilon} \tag{3.57}$$

in Planck units ($h = 2\pi\hbar = 2\pi$). The purely numerical integral has the value 3! multiplied by 1, $\pi^4/90$ and $\frac{7}{8}\pi^4/90$ for $\epsilon = 0, 1$ and $-1$ respectively, and we shall adopt 3!, absorbing any small discrepancy into $\mathcal{N}$. Then, from (3.53),

$$\rho = \frac{3\mathcal{N}}{\pi^2} T^4, \quad s = \frac{4\mathcal{N}}{\pi^2} T^3 \tag{3.58}$$

in terms of the local temperature $T$ given by (3.50).

Substituting (3.58) into (3.55) gives for the wall contribution to the total entropy

$$S_{wall} = \frac{4\mathcal{N}}{\pi^2} 4\pi r_1^2 T_\infty^3 \int_{r_1}^{r_1+\delta} \frac{dr}{f^2(r)}, \tag{3.59}$$

where $\delta$ is an arbitrary small length subject to $\Delta r \ll \delta \ll r_1$. It is useful to express this result in terms of the proper altitude $\alpha$ of the inner wall above the horizon $r = r_0$ of the exterior geometry (3.49). (Since (3.49) only applies for $r > r_1$, the physical space does not, of course, contain any horizon.) We assume that $f(r)$ has a (simple) zero for $r = r_0$, so we can write

$$f(r) \simeq 2\kappa_0 (r - r_0), \quad \kappa_0 = \frac{1}{2} f'(r_0) \neq 0 \quad (r \to r_0), \tag{3.60}$$

where $\kappa_0$ is the surface gravity. Then

$$\alpha = \int_{r_0}^{r_1} f^{-1/2} dr \quad \Rightarrow \quad \Delta r = \frac{1}{2}\kappa_0 \alpha^2, \tag{3.61}$$

and (3.59) can be written

$$S_{wall} = \frac{\mathcal{N}}{90\pi\alpha^2}\left(\frac{T_\infty}{\kappa_0/2\pi}\right)^3 \frac{1}{4}A \tag{3.62}$$

in Planck units, where $A = 4\pi r_1^2$ is the wall area.

Similarly, we find from (3.56) and (3.58) that thermal excitations near the wall contribute

$$\Delta M_{them,wall} = \frac{\mathcal{N}}{480\pi\alpha^2}\left(\frac{T_\infty}{\kappa_0/2\pi}\right)^3 AT_\infty \tag{3.63}$$

to the gravitational mass of the system.

The wall contribution to the free energy

$$F = \Delta M - T_\infty S \tag{3.64}$$

is

$$F_{wall} = -\frac{\mathcal{N}}{1440\pi\alpha^2}\left(\frac{T_\infty}{\kappa_0/2\pi}\right)^3 AT_\infty. \tag{3.65}$$

The entropy is recoverable from the free energy by the standard prescription

$$S_{wall} = -\partial F_{wall}/\partial T_\infty. \tag{3.66}$$

(Observe that this is an "off-shell" prescription [82, 32]: the geometrical quantities $A$, $\alpha$ and, in particular, the surface gravity $\kappa_0$ are kept fixed when the temperature is varied in (3.65).)

Following 'tHooft [35], we now introduce a crude cutoff to allow for quantum-gravity fluctuations by fixing the wall altitude $\alpha$ so that

$$S_{wall} = S_{BH}, \quad \text{when} \quad T_\infty = T_{BH}, \tag{3.67}$$

where the Bekenstein-Hawking entropy $S_{BH}$ and Hawking temperature $T_{BH}$ are defined to be the *purely geometrical* quantities in terms of the wall's area $A$ and redshifted acceleration (= surface gravity) $\kappa_0$. From (3.67) and (3.62), restoring conventional units for a moment, we find

$$\alpha = l_{pl}\sqrt{\mathcal{N}/90\pi}, \tag{3.68}$$

so that $\alpha$ is very near the Planck length if the effective number $\mathcal{N}$ of basic quantum fields in nature is on the order of 300.

It is significant and crucial that the normalization (3.68) is *universal*, depending only on fundamental physics, and independent of the mechanical and geometrical characteristics of the system.

With $\alpha$ fixed by (3.68), the wall's free energy (3.64) becomes

$$F_{wall} = -\frac{1}{16}\left(\frac{T_\infty}{T_{BH}}\right)^3 AT_\infty. \tag{3.69}$$

This "off-shell" formula expresses $F_{wall}$ in terms of three independent variables: the temperature $T_\infty$ and the geometrical characteristics $A$ and $T_{BH}$. From (3.69) we can obtain the wall entropy either from the thermodynamical Gibbs relation (3.66) (with $T_\infty$ set equal to $T_{BH}$ *after* differentiation), or from the Gibbs-Duhem formula (3.64) which is equivalent to the statistical-mechanical definition $S = -\mathbf{Tr}(\rho \ln \rho)$. Thus the distinction [82, 32] between "thermodynamical" and "statistical" entropies disappears in this formulation, because the geometrical and thermal variables are kept independent.

The wall's thermal mass-energy is given "on-shell" ($T_\infty = T_{BH}$) by

$$\Delta M_{therm,wall} = \frac{3}{16} A T_{BH} \tag{3.70}$$

according to (3.63) and (3.68). For a wall skirting a Schwarzschild horizon, so that $T_{BH} = (8\pi M)^{-1}$, this reduces to 'tHooft's result (3.38).

As already noted, thermal energy is not the only source of the wall's mass. Quantum fields outside the wall have as their ground state the Boulware state, which has a negative energy density growing to Planck levels near the wall. On shell, this very nearly cancels the thermal energy density (3.58); their sum is, in fact, the Hartle-Hawking value (cf. (3.45) and (3.48)):

$$(T^\nu_\mu)_{therm,T_\infty=T_{BH}} + (T^\nu_\mu)_B = (T^\nu_\mu)_{HH}, \tag{3.71}$$

which remains bounded near horizons, and integrates virtually to zero for a very thin layer near the wall. The total gravitational mass of the wall is thus, from (3.63) and (3.68),

$$\begin{aligned} (\Delta M)_{wall} &= (\Delta M)_{therm,wall} + (\Delta M)_{B,wall} \\ &= \frac{3}{16} A T_{BH} \left( (T_\infty / T_{BH})^4 - 1 \right), \end{aligned} \tag{3.72}$$

which vanishes on shell. For a central mass which is large in Planck units, there is no appreciable back-reaction of material near the wall on the background geometry (3.49).

We may conclude that many earlier concerns [35, 74, 75] were unnecessary: 'tHooft's brick wall model does provide a perfectly self-consistent description of a configuration which is indistinguishable from a black hole to outside observers, and which accounts for the Bekenstein-Hawking entropy purely as thermal entropy of quantum fields at the Hawking temperature (i.e. in the Hartle-Hawking state), providing one accepts the ad hoc but plausible ansatz (3.68) for a Planck-length cutoff near the horizon.

The model does, however, present us with a feature which is theoretically possible but appears strange and counterintuitive from a gravitational theorist's point of view. Although the wall is insubstantial (just like a horizon)—i.e., space there is practically a vacuum and the local curvature low—it is nevertheless the repository of all of the Bekenstein-Hawking entropy in the model.

It has been argued [39] that this is just what might be expected of black hole entropy in the entanglement picture. Entanglement will arise from virtual pair-creation in which one partner is "invisible" and the other "visible" (although only temporarily—nearly all get reflected back off the potential barrier). Such virtual pairs are all created very near the horizon. Thus, on this picture, the entanglement entropy (and its divergence) arises almost entirely from the strong correlation between nearby field variables on the two sides of the partition, an effect already present in flat space [83].

An alternative (but not necessarily incompatible) possibility is that the concentration of entropy at the wall is an artifact of the model or of the choice of Fock representation (based on a static observer's definition of positive frequency). The boundary condition of perfect reflectivity at the wall has no black hole counterpart. Moreover, one may well suspect that localization of entanglement entropy is not an entirely well-defined concept [47] or invariant under changes of the Fock representation.

### 3.2.3 The brick wall model reexamined

In the previous section, we have investigated the statistical mechanics of quantum fields in the region $r_1 < r < L$ of the spherical background (3.49) with the Dirichlet boundary condition at the boundaries. By using the particle description with the local temperature given by the Tolman's law, we have obtained the inner-wall contributions of the fields to entropy and thermal energy. When the former is set to be equal to the black hole entropy by fixing the cutoff $\alpha$ as (3.68), the later becomes comparable with the mass of the background geometry. After that, it has been shown that at the Hawking temperature the wall contribution to the thermal energy is exactly canceled by the negative energy of the Boulware state, assuming implicitly that the ground state of the model is the Boulware state and that the gravitational energy appearing in the Einstein equation is a sum of the renormalized energy of the Boulware state and the thermal energy of the fields.

In this section we shall show that these implicit assumptions do hold. In the following arguments it will also become clear how the local description used in the previous section is derived from the quantum field theory in curved spacetime, which is globally defined.

For simplicity, we consider a real scalar field described by the action

$$I = -\frac{1}{2}\int d^4x\sqrt{-g}\left[g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + m_\phi^2\phi^2\right]. \tag{3.73}$$

On the background given by (3.49), the action is reduced to

$$I = \int dtL, \tag{3.74}$$

with the Lagrangian $L$ given by

$$L = -\frac{1}{2}\int d^3x r^2\sqrt{\Omega}\left[-\frac{1}{f}(\partial_t\phi)^2 + f(\partial_r\phi)^2 + \frac{1}{r^2}\Omega^{ij}\partial_i\phi\partial_j\phi + m_\phi^2\phi^2\right]. \tag{3.75}$$

Here $x^i$ $(i = 1, 2)$ are coordinates on the 2-sphere. In order to make our system finite let us suppose that two mirror-like boundaries are placed at $r = r_1$ and $r = L$ $(r_1 < L)$, respectively, and investigate the scalar field in the region between the two boundaries. In the following arguments we quantize the scalar field with respect to the Killing time $t$. Hence, the ground state obtained below is the Boulware state. After the quantization, we investigate the statistical mechanics of the scalar field in the Boulware state. It will be shown that the resulting statistical mechanics is equivalent to the brick wall model.

Now let us proceed to the quantization procedure. First, the momentum conjugate to $\phi(r, x^i)$ is

$$\pi(r, x^i) = \frac{r^2\sqrt{\Omega}}{f}\partial_t\phi, \tag{3.76}$$

and the Hamiltonian is given by

$$H = \frac{1}{2}\int d^3x\left[\frac{f}{r^2\sqrt{\Omega}}\pi^2 + r^2\sqrt{\Omega}f\left(\partial_r\phi\right)^2 + \sqrt{\Omega}\Omega^{ij}\partial_i\phi\partial_j\phi + r^2\sqrt{\Omega}m_\phi^2\phi^2\right]. \tag{3.77}$$

Next, promote the field $\phi$ to an operator and expand it as

$$\phi(r, x^i) = \sum_{nlm}\frac{1}{\sqrt{2\omega_{nl}}}\left[a_{nlm}\varphi_{nl}(r)Y_{lm}(x^i)e^{-i\omega_{nl}t} + a_{nlm}^\dagger\varphi_{nl}(r)Y_{lm}(x^i)e^{i\omega_{nl}t}\right], \tag{3.78}$$

where $Y_{lm}(x^i)$ are real spherical harmonics defined by

$$\begin{aligned}\frac{1}{\sqrt{\Omega}}\partial_i\left(\sqrt{\Omega}\Omega^{ij}\partial_j Y_{lm}\right)+l(l+1)Y_{lm} &= 0,\\ \int Y_{lm}(x^i)Y_{l'm'}(x^i)\sqrt{\Omega(x^i)}d^2x &= \delta_{ll'}\delta_{mm'},\end{aligned}$$

and $\{\varphi_{nl}(r)\}$ $(n=1,2,\cdots)$ is a set of real functions defined below, which is complete with respect to the space of $L_2$-functions on the interval $r_1 \le r \le L$ for each $l$. The positive constant $\omega_{nl}$ is defined as the corresponding eigenvalue.

$$\begin{aligned}\frac{1}{r^2}\partial_r\left(r^2 f\partial_r\varphi_{nl}\right)-\left[\frac{l(l+1)}{r^2}+m_\phi^2\right]\varphi_{nl}+\frac{\omega_{nl}^2}{f}\varphi_{nl} &= 0, \qquad (3.79)\\ \varphi_{nl}(r_1)=\varphi_{nl}(L) &= 0,\\ \int_{r_1}^{L}\varphi_{nl}(r)\varphi_{n'l}(r)\frac{r^2}{f(r)}dr &= \delta_{nn'}.\end{aligned}$$

The corresponding expansion of the operator $\pi(r,x^i)$ is then:

$$\begin{aligned}\pi(r,x^i) &= -i\frac{r^2\sqrt{\Omega(x^i)}}{f(r)}\sum_{nlm}\sqrt{\frac{\omega_{nl}}{2}}\\ &\quad\times\left[a_{nlm}\varphi_{nl}(r)Y_{lm}(x^i)e^{-i\omega_n t}-a_{nlm}^\dagger\varphi_{nl}(r)Y_{lm}(x^i)e^{i\omega_{nl}t}\right]. \qquad (3.80)\end{aligned}$$

Hence, the usual equal-time commutation relation

$$\begin{aligned}\left[\phi(r,x^i),\pi(r',x'^i)\right] &= i\delta(r-r')\delta^2(x^i-x'^i),\\ \left[\phi(r,x^i),\phi(r',x'^i)\right] &= \left[\pi(r,x^i),\pi(r',x'^i)\right]=0 \qquad (3.81)\end{aligned}$$

becomes

$$\begin{aligned}\left[a_{nlm},a_{n'l'm'}^\dagger\right] &= \delta_{nn'}\delta_{ll'}\delta_{mm'},\\ \left[a_{nlm},a_{n'l'm'}\right] &= 0,\\ \left[a_{nlm}^\dagger,a_{n'l'm'}^\dagger\right] &= 0. \qquad (3.82)\end{aligned}$$

The normal-ordered Hamiltonian is given by

$$:H:=\sum_{nlm}\omega_{nl}a_{nlm}^\dagger a_{nlm}. \qquad (3.83)$$

Thus, the Boulware state $|B\rangle$, which is defined by

$$a_{nlm}|B\rangle=0 \qquad (3.84)$$

for $^\forall(n,l,m)$, is an eigenstate of the normal-ordered Hamiltonian with the eigenvalue zero. The Hilbert space of all quantum states of the scalar field is constructed as a symmetric Fock space on the Boulware state, and the complete basis $\{|\{N_{nlm}\}\rangle\}$ $(N_{nlm}=0,1,2,\cdots)$ is defined by

$$|\{N_{nlm}\}\rangle=\prod_{nlm}\frac{1}{\sqrt{N_{nlm}!}}\left(a_{nlm}^\dagger\right)^{N_{nlm}}|B\rangle, \qquad (3.85)$$

and each member of the basis is an eigenstate of the normal-ordered Hamiltonian:

$$:H:|\{N_{nlm}\}\rangle=\left(\sum_{nlm}\omega_{nl}N_{nlm}\right)|\{N_{nlm}\}\rangle. \qquad (3.86)$$

Now we shall investigate the statistical mechanics of the quantized scalar field. The free energy $F$ is given by

$$e^{-\beta_\infty F} \equiv \mathbf{Tr}\left[e^{-\beta_\infty :H:}\right] = \prod_{nlm} \frac{1}{1-e^{-\beta_\infty \omega_{nl}}}, \tag{3.87}$$

where $\beta_\infty = T_\infty^{-1}$ is inverse temperature. For explicit calculation of the free energy we adopt the WKB approximation. First, we rewrite the mode function $\varphi_{nl}(r)$ as

$$\varphi_{nl}(r) = \psi_{nl}(r)e^{-ikr}, \tag{3.88}$$

and suppose that the prefactor $\psi_{nl}(r)$ varies very slowly:

$$\left|\frac{\partial_r \psi_{nl}}{\psi_{nl}}\right| \ll |k|, \qquad \left|\frac{\partial_r^2 \psi_{nl}}{\psi_{nl}}\right| \ll |k|^2. \tag{3.89}$$

Thence, assuming that

$$\left|\frac{\partial_r (r^2 f)}{r^2 f}\right| \ll |k|, \tag{3.90}$$

the field equation (3.79) of the mode function is reduced to

$$k^2 = k^2(l,\omega_{nl}) \equiv \frac{1}{f}\left[\frac{\omega_{nl}^2}{f} - \frac{l(l+1)}{r^2} - m_\phi^2\right]. \tag{3.91}$$

Here we mention that the slowly varying condition (3.89) can be derived from the condition (3.90) and viceversa. The number of modes with frequency less than $\omega$ is given approximately by

$$\tilde{g}(\omega) = \int \nu(l,\omega)(2l+1)dl, \tag{3.92}$$

where $\nu(l,\omega)$ is the number of nodes in the mode with $(l,\omega)$:

$$\nu(l,\omega) = \frac{1}{\pi}\int_{r_1}^{L} \sqrt{k^2(l,\omega)}dr. \tag{3.93}$$

Here it is understood that the integration with respect to $r$ and $l$ is taken over those values which satisfy $r_1 \le r \le L$ and $k^2(l,\omega) \ge 0$. Thus, when

$$\left|\frac{\partial_r (r^2 f)}{r^2 f}\right| \ll \frac{1}{f\beta_\infty}$$

is satisfied, the free energy is given approximately by

$$F \simeq \frac{1}{\beta_\infty}\int_0^\infty \ln\left(1-e^{-\beta_\infty \omega}\right)\frac{d\tilde{g}(\omega)}{d\omega}d\omega = \int_{r_1}^{L} \tilde{F}(r)4\pi r^2 dr, \tag{3.94}$$

where the 'free energy density' $\tilde{F}(r)$ is defined by

$$\tilde{F}(r) \equiv \frac{1}{\beta(r)}\int_0^\infty \ln\left(1-e^{-\beta(r)E}\right)\frac{4\pi p^2 dp}{(2\pi)^3}. \tag{3.95}$$

Here the 'local inverse temperature' $\beta(r)$ is defined by the Tolman's law

$$\beta(r) = f^{1/2}(r)\beta_\infty, \tag{3.96}$$

and $E$ is defined by $E = \sqrt{p^2 + m_\phi^2}$. Hence the total energy $U$ (equal to $\Delta M_{therm}$ given by (3.56)) and entropy $S$ are calculated as

$$\begin{aligned} U &\equiv \mathbf{Tr}\left[e^{\beta_\infty(F-:H:)} : H :\right] = \frac{\partial}{\partial\beta_\infty}(\beta_\infty F) = \int_{r_1}^{L} \rho(r) 4\pi r^2 dr, \qquad (3.97) \\ S &\equiv -\mathbf{Tr}\left[e^{\beta_\infty(F-:H:)} \ln e^{\beta_\infty(F-:H:)}\right] = \beta_\infty^2 \frac{\partial}{\partial\beta_\infty} F \\ &= \int_{r_1}^{L} s(r) 4\pi r^2 dr/\sqrt{f(r)}, \qquad (3.98) \end{aligned}$$

where the 'density' $\rho(r)$ and the 'entropy density' $s(r)$ are defined by

$$\begin{aligned} \rho(r) &\equiv \frac{\partial}{\partial\beta(r)}(\beta(r)\tilde{F}(r)) = \int_0^\infty \frac{E}{e^{\beta(r)E}-1}\frac{4\pi p^2 dp}{(2\pi)^3}, \\ s(r) &\equiv \beta^2(r)\frac{\partial}{\partial\beta(r)}\tilde{F}(r) = \beta(r)\left(\rho(r) + P(r)\right), \end{aligned} \qquad (3.99)$$

where the 'pressure' $P(r)$ is defined by [1]

$$P(r) \equiv -\tilde{F}(r) = \frac{1}{3}\int_0^\infty \frac{p^2/E}{e^{\beta(r)E}-1}\frac{4\pi p^2 dp}{(2\pi)^3}. \qquad (3.100)$$

These expressions are exactly same as expressions (3.53) for the local quantities in the statistical mechanics of gas of particles.

Thus, we have shown that the local description of the statistical mechanics used in subsection 3.2.2 is equivalent to that of the quantized field in the curved background, which is defined globally, and whose ground state is the Boulware state.

The stress energy tensor of the minimally coupled scalar field is given by

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta I}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\left(g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + m_\phi^2\phi^2\right). \qquad (3.101)$$

In particular, the $(tt)$-component is

$$T_t^t = -\frac{1}{2}\left[\frac{1}{f}(\partial_t\phi)^2 + f(\partial_r\phi)^2 + \frac{1}{r^2}\Omega^{ij}\partial_i\phi\partial_j\phi + m_\phi^2\phi^2\right]. \qquad (3.102)$$

Hence, the contribution $\Delta M$ of the scalar field to the mass of the background geometry is equal to the Hamiltonian of the field:

$$\Delta M \equiv -\int_{r_1}^{L} T_t^t 4\pi r^2 dr = H, \qquad (3.103)$$

where $H$ is given by (3.77). When we consider the statistical mechanics of the hot quantized system, contributions of both vacuum polarization and thermal excitations must be taken into account. Thus, the contribution to the mass is given by

$$\langle\Delta M\rangle = \mathbf{Tr}\left[e^{\beta_\infty(F-:H:)}\Delta M^{(ren)}\right], \qquad (3.104)$$

where $\Delta M^{(ren)}$ is an operator defined by the expression (3.103) with $T_t^t$ replaced by the renormalized stress energy tensor $T^{(ren)}{}_t^t$. From (3.103), it is easy to show that

$$\Delta M^{(ren)} =: H : + \Delta M_B, \qquad (3.105)$$

[1] To obtain the last expression of $P(r)$ we performed an integration by parts.

where : $H$ : is the normal-ordered Hamiltonian given by (3.83) and $\Delta M_B$ is the zero-point energy of the Boulware state defined by

$$\Delta M_B = -\int_{r_1}^{L} \langle B|T^{(ren)}{}_t^t|B\rangle 4\pi r^2 dr. \tag{3.106}$$

Hence, $\langle \Delta M \rangle$ can be decomposed into the contribution of the thermal excitations and the contribution from the zero-point energy:

$$\langle \Delta M \rangle = U + \Delta M_B, \tag{3.107}$$

where $U$ is given by (3.97) and equal to $\Delta M_{therm}$ defined in (3.56).

Finally, we have shown that the gravitational mass appearing in the Einstein equation is the sum of the energy of the thermal excitation and the mass-energy of the Boulware state. Therefore, as shown in subsection 3.2.2, the wall contribution to the total gravitational mass is zero on shell ($T_\infty = T_{BH}$) and the backreaction can be neglected. Here, we mention that the corresponding thermal state on shell is called a topped-up Boulware state [36], and can be considered as a generalization to spacetimes not necessarily containing a black hole of the Hartle-Hawking state [77].

### 3.2.4 Complementarity

Attempts to provide a microscopic explanation of the Bekenstein-Hawking entropy $S_{BH}$ initially stemmed from two quite different directions.

Gibbons and Hawking [28] took the view that $S_{BH}$ is of topological origin, depending crucially on the presence of a horizon. They showed that $S_{BH}$ emerges as a boundary contribution to the geometrical part of the Euclidean action. (A non-extremal horizon is represented by a regular point in the Euclidean sector, so the presence of a horizon corresponds to the *absence* of an inner boundary in this sector.)

'tHooft [35] sought the origin of $S_{BH}$ in the thermal entropy of ambient quantum fields raised to the Hawking temperature. He derived an expression which is indeed proportional to the area, but with a diverging coefficient which has to be regulated by interposing a "brick wall" just above the gravitational radius and adjusting its altitude by hand to reproduce $S_{BH}$ with the correct coefficient.

In addition, the brick wall model appears to have several problematical features—large thermal energy densities near the wall, producing a substantial mass correction from thermal excitations—which have raised questions about its self-consistency as a model in which gravitational back-reaction is neglected.

We have shown that such caveats are seen to be unfounded once the ground state of the model is identified correctly. Since there are no horizons above the brick wall, the ground state is the Boulware state, whose negative energy almost exactly neutralizes the positive energy of the thermal excitations. 'tHooft's model is thus a perfectly self-consistent description of a configuration which to outside observers appears as a black hole but does not actually contain horizons.

It is a fairly widely held opinion (e.g. [83, 84]) that the entropy contributed by thermal excitations or entanglement is a one-loop correction to the zero-loop (or "classical") Gibbons-Hawking contribution. The viewpoint advocated in this section appears (at least superficially) quite different. We view these two entropy sources—(a) brick wall, no horizon, strong thermal excitations near the wall, Boulware ground state; and (b) black hole, horizon, weak (Hartle-Hawking) stress - energy near the horizon, Hartle-Hawking ground state—as ultimately equivalent but mutually exclusive (complementary in the sense of Bohr) descriptions of what is externally virtually the same physical situation. The near-vacuum experienced

by free-falling observers near the horizon is eccentrically but defensibly explainable, in terms of the description (a), as a delicate cancellation between a large thermal energy and an equally large and negative ground-state energy—just as the Minkowski vacuum is explainable to a uniformly accelerated observer as a thermal excitation above his negative-energy (Rindler) ground state. (This corresponds to setting $f(r) = r$ in the $(1 + 1)$-dimensional example treated in subsection 3.2.1.)

That the entropy of thermal excitations can single-handedly account for $S_{BH}$ without cutoffs or other *ad hoc* adjustments can be shown by a thermodynamical argument [36]. One considers the reversible quasi-static contraction of a massive thin spherical shell toward its gravitational radius. The exterior ground state is the Boulware state, whose stress-energy diverges to large negative values in the limit. To neutralize the resulting backreaction, the exterior is filled with thermal radiation to produce a "topped-up" Boulware state (TUB) whose temperature equals the acceleration temperature at the shell's radius. To maintain thermal equilibrium (and hence applicability of the first law) the shell itself must be raised to the same temperature. The first law of thermodynamics then shows that the shell's entropy approaches $S_{BH}$ (in the non-extremal case) for essentially arbitrary equations of state. Thus, the (shell + TUB) configuration passes smoothly to a black hole + Hartle-Hawking state in the limit.

It thus appears that one has two complementary descriptions, (a) and (b), of physics near an event horizon, corresponding to different Fock representations, i.e., different definitions of positive frequency and ground state. The Bogoliubov transformation that links these representations is known [85]. However, because of the infinite number of field modes, the two ground states are unitarily inequivalent [86]. This signals some kind of phase transition (formation of a condensate) in the passage between description (a), which explains $S_{BH}$ as a thermal effect, and description (b), which explains it as geometry. We know that a condensation actually does occur at this point; it is more usually called gravitational collapse.

It will be interesting to explore the deeper implications of these connections.

## 3.3 Entanglement entropy and thermodynamics

As explained several times, entanglement entropy is often speculated as a strong candidate for the origin of the black-hole entropy. To judge whether this speculation is true or not, it is effective to investigate the whole structure of thermodynamics obtained from the entanglement entropy, rather than just to examine the apparent structure of the entropy alone or to compare it with that of the black hole entropy. It is because entropy acquires a physical significance only when it is related to the energy and the temperature of a system. From this point of view, we construct a 'entanglement thermodynamics' by introducing an entanglement energy and compare it with the black-hole thermodynamics.

Our strategy of the 'construction of entanglement thermodynamics' is as follows. (See chapter 1 for the 'construction of black hole thermodynamics' for Schwarzschild black holes.)

1. First we give concepts and definitions of 'entanglement energy'. (We follow the usual definition for entanglement entropy.)

2. Next we calculate entanglement entropy and entanglement energy for tractable models.

3. Finally, we obtain 'entanglement temperature' $T_{ent}$ by assuming the following relation analogous to the first law of thermodynamics.

$$\delta E_{ent} = T_{ent} \delta S_{ent}, \tag{3.108}$$

where $S_{ent}$ and $E_{ent}$ denote entanglement entropy and energy, respectively. We call this relation the first law of entanglement thermodynamics.

This section is organized as follows. In subsection 3.3.1 we review the concept of the entanglement entropy. In subsection 3.3.2 we propose four definitions of entanglement energy and present general formulas for calculating the energy. In subsection 3.3.3 explicit evaluations of entanglement entropy and energy are performed for some tractable models with the help of the formulas prepared in subsection 3.3.2. In subsection 3.3.4 we construct entanglement thermodynamics and compare it with the black hole thermodynamics.

### 3.3.1 Entanglement entropy

In this subsection we review the definition and basic properties of the entanglement entropy.

#### Definition of entanglement entropy

Let $\mathcal{F}$ be a Hilbert space constructed from two Hilbert spaces $\mathcal{F}_1$ and $\mathcal{F}$ as

$$\mathcal{F} = \mathcal{F}_1 \bar{\otimes} \mathcal{F}_2, \tag{3.109}$$

where $\bar{\otimes}$ denotes a tensor product followed by a suitable completion. We call an element $u \in \mathcal{F}$ *prime* if $u$ can be written as $u = v \otimes w$ with $v \in \mathcal{F}_1$ and $w \in \mathcal{F}_2$. For example, $u = v_1 \otimes w_1 + 2v_1 \otimes w_2 + v_2 \otimes w_1 + 2v_2 \otimes w_2$ is prime since $u$ can be represented as $u = (v_1 + v_2) \otimes (w_1 + 2w_2)$. On the other hand $u = v_1 \otimes w_1 + v_2 \otimes w_2$ is not prime if neither $v_1$ and $v_2$ nor $w_1$ and $w_2$ are linearly dependent. The entanglement entropy $S_{ent} : \mathcal{F} \to \mathbf{R}_+ = \{\text{non-negative real numbers}\}$ defined below can be regarded as a measure of the non-prime nature of an element of $\mathcal{F} = \mathcal{F}_1 \bar{\otimes} \mathcal{F}_2$.

First of all, from an element $u$ of $\mathcal{F}$ with unit norm we construct an operator $\rho$ ('density operator') by

$$\rho v = (u, v)u \quad {}^\forall v \in \mathcal{F}, \tag{3.110}$$

where $(u, v)$ is the inner product which is antilinear with respect to $u$. In this context $\rho$ represents a 'pure state'.

From $\rho$ we define another operator ('reduced density operator') $\rho_2$ by

$$\rho_2 y = \sum_{i,j} f_j (e_i \otimes f_j, \rho e_i \otimes y) \quad {}^\forall y \in \mathcal{F}_2, \tag{3.111}$$

where $\{e_i\}$ and $\{f_j\}$ are orthonormal bases of $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. Note that

$$\mathrm{Tr}_2 (\rho_2 A) = \mathrm{Tr} [\rho (1 \otimes A)] \tag{3.112}$$

for an arbitrary bounded operator $A$ on $\mathcal{F}_2$.

Finally we define the entanglement entropy with respect to $\rho$ as

$$S_{ent} [\rho] \equiv -k_B \mathrm{Tr}_2 [\rho_2 \ln \rho_2] . \tag{3.113}$$

We can totally exchange the roles played by $\mathcal{F}_1$ and $\mathcal{F}_2$ in Eq.(3.111) and Eq.(3.113). The entanglement entropy is so defined as to be invariant under the exchange of $\mathcal{F}_1$ and $\mathcal{F}_2$ when $\rho$ corresponds to a pure state, i.e., when $\rho$ is given by Eq.(3.110). (See *Appendix* A.4 for the proof of this property.)

**A simple example**

As a simple example, let us consider spin states for a system consisting of an electron and a proton. We take $\mathcal{F}_1 = \{|\uparrow\rangle_e, |\downarrow\rangle_e\}$ for an electron and $\mathcal{F}_2 = \{|\uparrow\rangle_p, |\downarrow\rangle_p\}$ for a proton, where '$\uparrow$' is for 'up', while '$\downarrow$' is for 'down'. Then $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ is spanned by

$$\{|\uparrow\rangle_e \otimes |\uparrow\rangle_p, |\uparrow\rangle_e \otimes |\downarrow\rangle_p, |\downarrow\rangle_e \otimes |\uparrow\rangle_p, |\downarrow\rangle_e \otimes |\downarrow\rangle_p\}.$$

Now let us consider a state

$$\begin{aligned} |\phi\rangle &= (\alpha|\uparrow\rangle_e + \beta|\downarrow\rangle_e) \otimes (\gamma|\uparrow\rangle_p + \delta|\downarrow\rangle_p), \\ & |\alpha|^2 + |\beta|^2 = |\gamma|^2 + |\delta|^2 = 1, \end{aligned}$$

which is clearly a prime state. According to Eq.(3.111), we then get

$$\rho_e = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}.$$

Here '$e$' is for 'electron'. By a suitable diagonalization of this matrix, it is easy to see that $S_{ent} = 0$. We can exchange the roles between 'electron' and 'proton': we then get

$$\rho_p = \begin{pmatrix} |\gamma|^2 & \gamma\delta^* \\ \gamma^*\delta & |\delta|^2 \end{pmatrix},$$

('$p$' is for 'proton') which again leads to $S_{ent} = 0$.

On the contrary, an $s$-state

$$\begin{aligned} |\phi'\rangle &= \alpha|\uparrow\rangle_e \otimes |\downarrow\rangle_p + \beta|\downarrow\rangle_e \otimes |\uparrow\rangle_p, \\ & |\alpha|^2 + |\beta|^2 = 1, \alpha\beta \neq 0 \end{aligned}$$

is not a prime state. For this state the reduced density operators are given by

$$\rho_e = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}, \rho_p = \begin{pmatrix} |\beta|^2 & 0 \\ 0 & |\alpha|^2 \end{pmatrix}.$$

Therefore we get $S_{ent} = -k_B(|\alpha|^2 \ln|\alpha|^2 + |\beta|^2 \ln|\beta|^2) > 0$.

**Formula of entanglement entropy**

Let us consider a system of coupled harmonic oscillators $\{q^A\}$ $(A = 1, \cdots, n_{tot})$ described by the Lagrangian,

$$L = \frac{a}{2}\delta_{AB}\dot{q}^A\dot{q}^B - \frac{1}{2}V_{AB}q^Aq^B. \tag{3.114}$$

Here $\delta_{AB}$ is Kronecker's delta symbol[2]; $V$ is a real-symmetric, positive-definite matrix which does not depend on $\{q^A\}$. We have introduced $a(>0)$ as a fundamental length characterizing the system.[3] The corresponding Hamiltonian becomes

$$H_{tot} = \frac{1}{2a}\delta^{AB}p_Ap_B + \frac{1}{2}V_{AB}q^Aq^B, \tag{3.115}$$

where $p_A = a\delta_{AB}\dot{q}^B$ is the canonical momentum conjugate to $q^A$.

[2] From now on, we choose the units $\hbar = c = 1$ and apply Einstein's summation convention unless otherwise stated.

[3] Thus $\{q^A\}$ are treated as dimension-free quantities in the present units.

Firstly we calculate the wave function $\langle\{q^A\}|0\rangle$ of the ground state $|0\rangle$. Note that Eq.(3.115) can be written as

$$H_{tot} = \frac{1}{2a}\delta^{AB}\left(p_A + iW_{AC}q^C\right)\left(p_B - iW_{BD}q^D\right) + \frac{1}{2a}\mathrm{Tr}W \tag{3.116}$$

by using the commutation relation $\left[q^A, p_B\right] = i\delta^A_B$. Here $W$ is a symmetric matrix satisfying $(W^2)_{AB} = aV_{AB}$. The ambiguity in sign is fixed by requiring $W$ to be positive definite. Thus,

$$W = \sqrt{aV}. \tag{3.117}$$

Now $\langle\{q^A\}|0\rangle$ is given as a solution to

$$\left(\frac{\partial}{\partial q^A} + W_{AB}q^B\right)\langle\{q^A\}|0\rangle = 0, \tag{3.118}$$

since $p_A$ is expressed as $-i\frac{\partial}{\partial q^A}$. The solution is

$$\langle\{q^A\}|0\rangle = \left(\det\frac{W}{\pi}\right)^{1/4}\exp\left(-\frac{1}{2}W_{AB}q^Aq^B\right), \tag{3.119}$$

which is normalized with respect to the standard Lebesgue measure $dq^1\cdots dq^{n_{tot}}$. The density matrix $\rho_0$ corresponding to this ground state is represented as

$$\begin{aligned}\langle\{q^A\}|\rho_0|\{q'^B\}\rangle &= \langle\{q^A\}|0\rangle\langle 0|\{q'^B\}\rangle \\ &= \left(\det\left(\frac{W}{\pi}\right)\right)^{1/2} \\ &\quad\times\exp\left[-\frac{1}{2}W_{AB}\left(q^Aq^B + q'^Aq'^B\right)\right].\end{aligned} \tag{3.120}$$

Now we split $\{q^A\}$ into two subsystems, $\{q^a\}$ $(a = 1,\cdots,n_B)$ and $\{q^\alpha\}$ $(\alpha = n_B+1,\cdots,n_{tot})$. (We assign the labels '1' and '2' to the former and the latter subsystems, respectively.) Then we obtain the reduced density matrix associated with the subsystem 2 (the subsystem 1), by taking the partial trace of $\rho_0$ w.r.t. the subsystem 1 (the subsystem 2):

$$\begin{aligned}\langle\{q^\alpha\}|\rho_2|\{q'^\beta\}\rangle &= \int\prod_{c=1}^{n}dq^c\langle\{q^a,q^\alpha\}|\rho_0|\{q^b,q'^\beta\}\rangle \\ &= \left(\det\frac{D'}{\pi}\right)^{1/2}\exp\left[-\frac{1}{2}D'_{\alpha\beta}\left(q^\alpha q^\beta + q'^\alpha q'^\beta\right)\right] \\ &\quad\times\exp\left[-\frac{1}{4}\left(B^TA^{-1}B\right)_{\alpha\beta}(q-q')^\alpha(q-q')^\beta\right]\end{aligned} \tag{3.121}$$

and

$$\begin{aligned}\langle\{q^a\}|\rho_1|\{q'^b\}\rangle &= \int\prod_{\gamma=n+1}^{N}dq^\gamma\langle\{q^a,q^\alpha\}|\rho_0|\{q'^b,q^\beta\}\rangle \\ &= \left(\det\frac{A'}{\pi}\right)^{1/2}\exp\left[-\frac{1}{2}A'_{ab}\left(q^aq^b + q'^aq'^b\right)\right] \\ &\quad\times\exp\left[-\frac{1}{4}\left(BD^{-1}B^T\right)_{ab}(q-q')^a(q-q')^b\right],\end{aligned} \tag{3.122}$$

where $A$, $B$, $D$, $A'$ and $D'$ are defined by

$$\begin{aligned}(W_{AB}) &= \begin{pmatrix} A_{ab} & B_{a\beta} \\ (B^T)_{\alpha b} & D_{\alpha\beta} \end{pmatrix} \quad , \\ A' &= A - BD^{-1}B^T \quad , \\ D' &= D - B^T A^{-1} B \quad . \end{aligned} \tag{3.123}$$

(The superscript $T$ denotes transposition.) Note that $A^T = A$ and $D^T = D$. Here we mention that, if we define $\tilde{A}$, $\tilde{B}$ and $\tilde{D}$ by

$$W^{-1} = \begin{pmatrix} \tilde{A}^{ab} & \tilde{B}^{a\beta} \\ (\tilde{B}^T)^{\alpha b} & \tilde{D}^{\alpha\beta} \end{pmatrix}, \tag{3.124}$$

then

$$\begin{aligned}\tilde{A} &= {A'}^{-1}, \\ \tilde{D} &= {D'}^{-1}, \\ \tilde{B} &= -{A'}^{-1} B D^{-1} = -A^{-1} B {D'}^{-1}. \end{aligned} \tag{3.125}$$

Now the entanglement entropy $S_{ent} := -\mathrm{Tr}\rho_2 \ln \rho_2$ is given as follows[37, 38]. Let $\{\Lambda_i\}$ $(i = 1, \cdots, N - n_B)$ be the eigenvalues of a positive definite symmetric matrix [4] $\Lambda$,

$$\Lambda := \tilde{D}^{1/2} B^T A^{-1} B \tilde{D}^{1/2}. \tag{3.126}$$

Then it is easily shown that entanglement entropy is given by

$$\begin{aligned} S_{ent} &= \sum_{i=1}^{N-n_B} S_i, \\ S_i &= -\frac{\mu_i}{1-\mu_i} \ln \mu_i - \ln(1-\mu_i), \end{aligned} \tag{3.127}$$

where $\mu_i := \Lambda_i^{-1} \left(\sqrt{1+\Lambda_i} - 1\right)^2$. (Note that $0 < \mu_i < 1$.)

**Relevance to black hole entropy**

In the case of black-hole physics, the presence of the event horizon causes a natural decomposition of a Hilbert space $\mathcal{F}$ of all states of matter fields to a tensor product of the state spaces inside and outside a black hole as Eq. (3.109). For example, let us take a scalar field. We can suppose that its one-particle Hilbert space $\mathcal{H}$ is decomposed as

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2, \tag{3.128}$$

where $\mathcal{H}_1$ is a space of mode functions with supports inside the horizon and $\mathcal{H}_2$ is a space of mode functions with supports outside the horizon. Then we can construct new Hilbert spaces ('Fock spaces') $\mathcal{F}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ from $\mathcal{H}$, $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, as

$$\begin{aligned} \mathcal{F} &\equiv \mathrm{C} \oplus \mathcal{H} \oplus (\mathcal{H} \bar{\otimes} \mathcal{H})_{sym} \oplus \cdots, \\ \mathcal{F}_1 &\equiv \mathrm{C} \oplus \mathcal{H}_1 \oplus (\mathcal{H}_1 \bar{\otimes} \mathcal{H}_1)_{sym} \oplus \cdots, \\ \mathcal{F}_2 &\equiv \mathrm{C} \oplus \mathcal{H}_2 \oplus (\mathcal{H}_2 \bar{\otimes} \mathcal{H}_2)_{sym} \oplus \cdots, \end{aligned} \tag{3.129}$$

[4] The corresponding expression in ref.[37] ("$\Lambda^a{}_b := (M^{-1})^{ac} N_{cb}$") reads $\Lambda = \tilde{D} B^T A^{-1} B$ in the present notation. This definition does not give a symmetric matrix and should be replaced by "$\Lambda^{ab} := (M^{-1/2})^{ac} N_{cd} (M^{-1/2})^{db}$", namely Eq.(3.126).

where $(\cdots)_{sym}$ denotes the symmetrization. Now these three Hilbert spaces satisfy the relation (3.109). Hence the entanglement entropy $S_{ent}$ is defined by the procedure given at the beginning of this subsection (Eqs.(3.110)-(3.113)) for each state in $\mathcal{F}$.

The entanglement entropy $S_{ent}$ originates from a tensor product structure of the Hilbert space as Eq.(3.109), which is caused by the existence of the boundary between two regions (the event horizon) through Eq.(3.128). Furthermore the symmetric property of $S_{ent}$ between $\mathcal{F}_1$ and $\mathcal{F}_2$ mentioned before also suggests that $S_{ent}$ is related with a boundary between two regions. In fact $S_{ent}$ turns out to be proportional to the area of such a boundary (a model for the event horizon) in simple models discussed below. In view of the Bekenstein-Hawking formula (1.1), thus, the entanglement entropy has a nature similar to the black hole entropy.

The relevance of the entanglement entropy to the black hole entropy is also suggested by the following observation. Let us consider a free scalar field on a background geometry describing a gravitational collapse to a black hole. We compare the black hole entropy and the entanglement entropy for this system. We begin with the black hole entropy. In the initial region of the spacetime, there is no horizon and the entropy around this region can be regarded as zero. In the final region, on the other hand, there is an event horizon so that the black-hole possesses non-zero entropy. As for the entanglement entropy, the existence of the event horizon naturally divides the Hilbert space $\mathcal{F}$ of all states of the scalar field into $\mathcal{F}_1 \bar{\otimes} \mathcal{F}_2$. Thus, the scalar field in some pure state possesses non-zero entanglement entropy. In this manner, we observe that the black-hole entropy and the entanglement entropy come from the same origin, i.e. the existence of the event horizon. This is the reason why the entanglement entropy is regarded as one of the potential candidates for the origin of the black-hole entropy.

**Simple models in Minkowski spacetime**

The relation between the entanglement entropy and the black hole entropy was analyzed in terms of simple tractable models by Bombelli, et. al. [37] and Srednicki [38]. They considered a free scalar field on a flat spacelike hypersurface embedded in a 4-dimensional Minkowski spacetime, and calculated the entanglement entropy for a division of the hypersurface into two regions with a common boundary $B$. Here the two regions and $B$ are, respectively, the models of the interior, the exterior of the black holes and the horizon. Ref.[37] chooses $B$ to be a 2-dimensional flat surface $\mathbf{R}^2$, and the matter state to be the ground state, showing that the resulting entanglement entropy becomes proportional to the area of $B$. Ref.[38] chooses $B$ to be a 2-sphere $S^2$ in $\mathrm{R}^3$, and chooses the two regions to be the interior and the exterior of the sphere. The matter state is chosen to be the ground state. Then it is shown that the resulting entanglement entropy is again proportional to the area of $B$.

Both of the results can be expressed as

$$S_{ent}[\rho_0] \simeq k_B \mathcal{N}_S \frac{A}{4\pi a^2}, \tag{3.130}$$

where $\rho_0$ is the ground-state density matrix, $A$ is area of the boundary, $a$ is a cutoff length, and $\mathcal{N}_S$ is a dimensionless numerical constant of order unity. In particular, $\mathcal{N}_S = 0.30$ for $B = S^2$ [38]. This coincides with the Bekenstein-Hawking formula if the cut-off length $a$ is chosen as

$$a = \sqrt{\frac{\mathcal{N}_S \hbar G}{\pi c^3}} = \sqrt{\frac{\mathcal{N}_S}{\pi}}\ l_{pl}, \tag{3.131}$$

where $l_{pl}$ is the Planck length. Here note that $a$ depends only on the Planck length.

Note that the cutoff length (3.131) is almost same as the cutoff length (3.68) for the brick wall model.

### 3.3.2 Entanglement energy

In this subsection we define entanglement energy to construct entanglement thermodynamics. We give four possible definitions of entanglement energy. The difference between them comes from the difference in the way to formulate the reduction of a system (caused by, for instance, the formation of an event horizon). In the first to third definitions we assume that some operators drop out from the set of all observables, while the state of a total system is regarded as unchanged. In the fourth definition, on the contrary, we assume that the state undergoes a change in the course of the reduction of the system (so that the density matrix of the system changes actually), while operators are regarded as unchanged. Since at present we cannot judge whether and which one of these treatments reflects the true process of reduction, the best way is to investigate all options. As we shall see in subsection 3.3.4, the universal behavior of the entanglement thermodynamics does not depend on the choice of the entanglement energy.

Let us consider a system described by a Fock space $\mathcal{F}$ constructed from a one-particle Hilbert space $\mathcal{H}$ in the previous section. Let $H_{tot}$ be a total Hamiltonian acting on $\mathcal{F}$. We assume that the Hamiltonian $H_{tot}$ is naturally decomposed as

$$H_{tot} = H_1 + H_2 + H_{int} \quad , \tag{3.132}$$

where $H_1$ and $H_2$ are parts acting on $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and $H_{int}$ is a part representing the interaction of two regions.

#### The first to third definitions of the entanglement energy

The first to third definitions of entanglement energy follow when we regard that the operators connecting the two subsystems drop out from the set of observables (when, for instance, an event horizon is formed), while the state of the system is regarded as unchanged. To be more precise, we assume that $H_1$ and $H_2$ remain to be observables but that $H_{int}$ is no longer an observable. In this case it is natural to define the entanglement energy by one of the following three.

(a) $E_{end} = \langle : H_1 : \rangle \equiv \mathrm{Tr}[: H_1 : \rho]$,

(b) $E_{end} = \langle : H_2 : \rangle \equiv \mathrm{Tr}[: H_2 : \rho]$,

(c) $E_{ent} = \langle : H_1 : \rangle + \langle : H_2 : \rangle$,

where $\rho$ is a density matrix of the total system and the two normal orderings mean to subtract the minimum eigenvalues of $H_1$ and $H_2$ respectively.

#### The forth definition of the entanglement energy

Next let us consider the case in which the total density operator $\rho$ actually changes to the product of reduced density operators of each subsystems, $\rho_1$ and $\rho_2$, (when, for instance, an event horizon is formed), while the observables remain unchanged.

In this case $\rho$ reduces to $\rho'$ given by

$$\rho' = \rho_1 \otimes \rho_2 \quad . \tag{3.133}$$

It is easy to see that the entropy associated with this density matrix becomes

$$-k_B \mathrm{Tr}\left[\rho' \ln \rho'\right] = S_{ent}[\rho] + S'_{ent}[\rho], \tag{3.134}$$

where $S_{ent}[\rho]$ and $S'_{ent}[\rho]$ are entanglement entropy obtained through $\rho_1$ and $\rho_2$, respectively. $S_{ent}[\rho]$ and $S'_{ent}[\rho]$ are identical if $\rho$ is a pure state (see the argument below Eq. (3.113)).

Since we are assuming that the observables do not change, we are led to the following forth definition of entanglement energy:

(d) $E_{ent} = \langle : H_{tot} : \rangle_{\rho'} \equiv \mathrm{Tr}\,[: H_{tot} : \rho']$,

where $: \ - \ :$ denotes the usual normal ordering (a subtraction of the ground state energy).

**Formula of $\langle : H_1 : \rangle$ for the ground state**

What we should do next is to give formulas of entanglement energy explicitly by choosing $\rho$ as the ground state $\rho_0$ of $H_{tot}$. In the next subsection we consider a free scalar field and discretize it with some spatial separation for regularization. Since the system thus obtained is equivalent to a set of harmonic oscillators, here, we give a formulas of entanglement energy for the ground state of coupled harmonic oscillators described by the Hamiltonian (3.115), splitting the total system as in the previous subsection.

First, we derive formulas for the entanglement energies corresponding to the definitions (a), (b) and (c).

Firstly we divide the Hamiltonian (3.115) into three terms as Eq.(3.132):

$$\begin{aligned}
H_1 &\equiv \frac{1}{2a}\delta^{ab}p_a p_b + \frac{1}{2}{V^{(1)}}_{ab}q^a q^b \\
&= \frac{1}{2a}\delta^{ab}\left(p_a + iw^{(1)}_{ac}q^c\right)\left(p_b - iw^{(1)}_{bd}q^d\right) + \frac{1}{2a}\mathrm{Tr}w^{(1)}, \\
H_2 &\equiv \frac{1}{2a}\delta^{\alpha\beta}p_\alpha p_\beta + \frac{1}{2}{V^{(2)}}_{\alpha\beta}q^\alpha q^\beta \\
&= \frac{1}{2a}\delta^{\alpha\beta}\left(p_\alpha + iw^{(2)}_{\alpha\gamma}q^\gamma\right)\left(p_\beta - iw^{(2)}_{\beta\delta}q^\delta\right) + \frac{1}{2a}\mathrm{Tr}w^{(2)}, \\
H_{int} &\equiv H_{tot} - H_1 - H_2 \\
&= {V_{int}}_{\ a\beta}q^a q^\beta,
\end{aligned} \tag{3.135}$$

where $V^{(1)}$, $V^{(2)}$ and $V_{int}$ are blocks in the matrix $V$ given by

$$(V_{AB}) = \begin{pmatrix} V^{(1)}_{ab} & (V_{int})_{a\beta} \\ ({V_{int}}^T)_{\alpha b} & V^{(2)}_{\alpha\beta} \end{pmatrix}, \tag{3.136}$$

and $w^{(1)}$ and $w^{(2)}$ are, respectively, the positive square-roots of $aV^{(1)}$ and $aV^{(2)}$. Although there exists freedom in the way of the division, the above division seems to be the most natural one. Here and throughout this section we adopt it.

By rescaling the variables $\left\{q^A\right\}$ as

$$\bar{q}^A := \delta^{AB}\left(W^{1/2}\right)_{BC}q^C,$$

the expression of the density matrix for the vacuum state Eq.(3.120) gets simplified as

$$\langle\left\{\bar{q}^A\right\}|\rho_0|\left\{\bar{q}'^B\right\}\rangle = \prod_{C=1}^{N}\pi^{-1/2}\exp\left[-\frac{1}{2}\left\{(\bar{q}^C)^2 + (\bar{q}'^C)^2\right\}\right],$$

and the normal ordered Hamiltonian $: H_1 :$ is represented as

$$\begin{aligned}
: H_1 : &= -\frac{1}{2a}\delta^{ab}\left(\frac{\partial}{\partial q^a} - w^{(1)}_{ac}q^c\right)\left(\frac{\partial}{\partial q^b} + w^{(1)}_{bd}q^d\right) \\
&= -\frac{1}{2a}U^{AB}\left(\frac{\partial}{\partial \bar{q}^A} - \bar{w}^{(1)}_{AC}\bar{q}^C\right)\left(\frac{\partial}{\partial \bar{q}^B} + \bar{w}^{(1)}_{BD}\bar{q}^D\right).
\end{aligned}$$

Here the matrices $U$ and $\bar{w}^{(1)}$ are defined as

$$\begin{aligned} U^{AB} &:= \delta^{AC}\left(W^{1/2}\right)_{Ca}\delta^{ab}\left(W^{1/2}\right)_{bD}\delta^{DB}, \\ \bar{w}^{(1)}_{AB} &:= \delta_{AC}\left(W^{-1/2}\right)^{Ca} w^{(1)}_{ab}\left(W^{-1/2}\right)^{bD}\delta_{DB}. \end{aligned}$$

Hence the matrix elements of $:H_1:\rho$ with respect to the basis $|\bar{q}^A\rangle$ are expressed as

$$\begin{aligned} &\langle\{\bar{q}^A\}\,|:H_1:\rho_0|\,\{\bar{q}'^B\}\rangle \\ &= \frac{1}{2a}\left\{\left[(\bar{w}^{(1)}+1)U(\bar{w}^{(1)}-1)\right]_{AB}\bar{q}^A\bar{q}^B + \mathrm{Tr}\left[U(1-\bar{w}^{(1)})\right]\right\} \\ &\times\prod_{C=1}^{N}\pi^{-1/2}\exp\left[-(\bar{q}^C)^2\right]. \end{aligned}$$

From this we obtain [5]

$$\begin{aligned} \langle:H_1:\rangle &= \int\left(\prod_{C=1}^{N}d\bar{q}^C\right)\langle\{\bar{q}^A\}\,|:H_1:\rho_0|\,\{\bar{q}'^B\}\rangle \\ &= \frac{1}{4a}\left[aV^{(1)}_{ab}(\tilde{A})^{ab} + A_{ab}\delta^{ab} - 2w^{(1)}_{ab}\delta^{ab}\right]. \end{aligned} \tag{3.137}$$

Similarly $\langle:H_2:\rangle$ is expressed as

$$\langle:H_2:\rangle = \frac{1}{4a}\left[aV^{(2)}_{\alpha\beta}(\tilde{D})^{\alpha\beta} + D_{\alpha\beta}\delta^{\alpha\beta} - 2w^{(2)}_{\alpha\beta}\delta^{\alpha\beta}\right], \tag{3.138}$$

where $w^{(2)}$ is the positive square-root of $aV^{(2)}$.

**Formula of $\langle:H_{tot}:\rangle_{\rho'}$ for the ground state**

By using the formulas (3.122) and (3.121), $\rho'$ defined by Eq.(3.133) is represented as

$$\begin{aligned} \langle\{q^A\}\,|\rho'|\,\{q'^B\}\rangle &= \left(\det\frac{M}{\pi}\right)^{1/2}\exp\left[-\frac{1}{2}M_{AB}\left(q^Aq^B + q'^Aq'^B\right)\right] \\ &\quad\times\exp\left[-\frac{1}{4}N_{AB}(q-q')^A(q-q')^B\right], \end{aligned} \tag{3.139}$$

where

$$\begin{aligned} (M_{AB}) &= \begin{pmatrix} A'_{ab} & 0 \\ 0 & D'_{\alpha\beta}\end{pmatrix}, \\ (N_{AB}) &= \begin{pmatrix} \left(BD^{-1}B^T\right)_{ab} & 0 \\ 0 & \left(B^TA^{-1}B\right)_{\alpha\beta}\end{pmatrix}. \end{aligned} \tag{3.140}$$

We can diagonalize $M$ and $N$ simultaneously by the following non-orthogonal transformation:

$$q^A \to \tilde{q}^A \equiv \left(\tilde{U}M^{1/2}\right)^A_{\ B}q^B, \tag{3.141}$$

where $\tilde{U}$ is a real orthogonal matrix satisfying

$$\begin{aligned} M^{-1/2}NM^{-1/2} &= \tilde{U}^T\lambda\tilde{U}, \\ \lambda &= \begin{pmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \ddots\end{pmatrix}. \end{aligned} \tag{3.142}$$

[5] See Eqs.(3.123) and (3.124) for the definitions of the matrices $A$, $\tilde{A}$, $D$ and $\tilde{D}$.

Now in terms of $\{\tilde{q}^A\}$, $H_{tot}$ is represented as

$$H_{tot} = -\frac{1}{2a}\left(\tilde{U}M\tilde{U}^T\right)^{AB}\left(\frac{\partial}{\partial\tilde{q}^A} - \tilde{W}_{AC}\tilde{q}^C\right)\left(\frac{\partial}{\partial\tilde{q}^B} + \tilde{W}_{BD}\tilde{q}^D\right) + \frac{1}{2a}\mathrm{Tr}W, \quad (3.143)$$

thus,

$$: H_{tot} := -\frac{1}{2a}\left(\tilde{U}M\tilde{U}^T\right)^{AB}\left(\frac{\partial}{\partial\tilde{q}^A} - \tilde{W}_{AC}\tilde{q}^C\right)\left(\frac{\partial}{\partial\tilde{q}^B} + \tilde{W}_{BD}\tilde{q}^D\right), \quad (3.144)$$

where

$$\tilde{W} \equiv \tilde{U}M^{-1/2}WM^{-1/2}\tilde{U}^T. \quad (3.145)$$

Hence the density matrix $\rho'$ is expressed in terms of $|\{\tilde{q}^A\}\rangle$ as[6]

$$\langle\{\tilde{q}^A\}|\rho'|\{\tilde{q}'^B\}\rangle = \prod_{C=1}^{N}\pi^{-1/2}\exp\left[-\frac{1}{2}\{(\tilde{q}^C)^2 + (\tilde{q}'^C)^2\} - \frac{1}{4}\lambda_C(\tilde{q}^C - \tilde{q}'^C)^2\right]. \quad (3.146)$$

This density matrix is normalized with respect to the measure $d\tilde{q}^1\cdots d\tilde{q}^N$.

Now it is easy to calculate the entanglement energy. First the matrix components of $:H_{tot}:\rho'$ with respect to $\{\tilde{q}^A\}$ are given by

$$\begin{aligned}&\langle\{\tilde{q}^A\}|:H_{tot}:\rho'|\{\tilde{q}^B\}\rangle \\ &= -\frac{1}{2a}\Big\{\left[\tilde{U}M\tilde{U}^T - l\tilde{U}M^{-1/2}VM^{-1/2}\tilde{U}^T\right]_{AB}\tilde{q}^A\tilde{q}^B \\ &+\mathrm{Tr}\left[W - N/2 - M\right]\Big\}\prod_{C=1}^{N}\pi^{-1/2}\exp\left[-(\tilde{q}^C)^2\right]. \end{aligned} \quad (3.147)$$

Hence the entanglement energy $\langle:H_{tot}:\rangle_{\rho'}$ is expressed as

$$\begin{aligned}\langle:H_{tot}:\rangle_{\rho'} &= \int(\prod_{C=1}^{N}d\tilde{q}^C)\langle\{\tilde{q}^C\}|:H_{tot}:\rho'|\{\tilde{q}^B\}\rangle \\ &= \frac{1}{4a}\mathrm{Tr}\left[aVM^{-1} + M + N - 2W\right]. \end{aligned} \quad (3.148)$$

Here we have used the formula $\int d\vec{x}\ \vec{x}\cdot\mathcal{A}\vec{x}\ \exp[-\vec{x}\cdot\vec{x}] = \frac{1}{2}\pi^{N/2}\mathrm{Tr}\mathcal{A}$, where $N$ is the dimension of $\vec{x}$. With the help of the identity $\mathrm{Tr}\left[M+N\right] = \mathrm{Tr}A + \mathrm{Tr}D = \mathrm{Tr}W$, we finally arrive at the following formula for $\langle:H_{tot}:\rangle_{\rho'}$

$$\begin{aligned}\langle:H_{tot}:\rangle_{\rho'} &= \frac{1}{4a}\mathrm{Tr}\left[aVM^{-1} - W\right] \\ &= \frac{1}{4}\mathrm{Tr}\left[V(M^{-1} - W^{-1})\right] \\ &= -\frac{1}{2}\mathrm{Tr}\left[V_{int}^T\tilde{B}\right]. \end{aligned} \quad (3.149)$$

**Alternative formula of $\langle:H_1:\rangle + \langle:H_2:\rangle$ for the ground state**

We can calculate $\langle:H_1:\rangle + \langle:H_2:\rangle$ for the ground state as a sum of Eqs. (3.137) and (3.138). Nonetheless, for a check of numerical calculations, it is useful to give an alternative formula of $\langle:H_1:\rangle + \langle:H_2:\rangle$.

[6] Einstein's summation convention is not applied to Eq.(3.146).

In terms of $\left\{\bar{q}^A\right\}$, the operator $:H_1:+:H_2:$ is written as

$$:H_1:+:H_2:=-\frac{1}{2a}\delta^{AC}\delta^{BD}W_{CD}\left(\frac{\partial}{\partial\bar{q}^A}-\bar{w}_{AE}\bar{q}^E\right)\left(\frac{\partial}{\partial\bar{q}^B}+\bar{w}_{BF}\bar{q}^F\right) \quad , \tag{3.150}$$

where $\bar{w}$ is defined by

$$\begin{aligned}\bar{w} &\equiv \delta_{AC}\left(W^{-1/2}wW^{-1/2}\right)^{CD}\delta_{DB} \quad , \\ (w_{AB}) &\equiv \begin{pmatrix} {w^{(1)}}_{ab} & 0 \\ 0 & {w^{(2)}}_{\alpha\beta} \end{pmatrix} \quad . \end{aligned} \tag{3.151}$$

From the expression we obtain

$$\begin{aligned}&\langle\left\{\bar{q}^A\right\}|(:H_1:+:H_2:)\rho_0|\left\{\bar{q}^B\right\}\rangle \\ &=\frac{1}{2a}\left\{\left[(\bar{w}+1)W(\bar{w}-1)\right]_{AB}\bar{q}^A\bar{q}^B-\mathrm{Tr}\left[W(\bar{w}-1)\right]\right\} \\ &\times\prod_{C=1}^{N}\pi^{-1/2}\exp\left[-(\bar{q}^C)^2\right].\end{aligned} \tag{3.152}$$

Hence we arrive at the following expression $\langle:H_1:\rangle+\langle:H_2:\rangle$ for $\rho_0$.

$$\begin{aligned}\langle:H_1:\rangle+\langle:H_2:\rangle &= \int(\prod_{C=1}^{N}d\bar{q}^C)\langle\left\{\bar{q}^A\right\}|(:H_1:+:H_2:)\rho_0|\left\{\bar{q}^B\right\}\rangle \\ &= \frac{1}{4a}\mathrm{Tr}\left[w^2W^{-1}-W\right]-\frac{1}{2a}\mathrm{Tr}\left[w-W\right].\end{aligned} \tag{3.153}$$

With the help of the relation $\mathrm{Tr}[w^2W^{-1}]=\mathrm{Tr}[aVM^{-1}]$ which follows from the definitions of $w$ and $M$, this formula is simplified as

$$\begin{aligned}\langle:H_1:\rangle+\langle:H_2:\rangle &= \frac{1}{4a}\mathrm{Tr}\left[aVM^{-1}-W\right]-\frac{1}{2a}\mathrm{Tr}\left[w-W\right] \\ &= \langle:H_{tot}:\rangle_{\rho'}-\frac{1}{2a}\mathrm{Tr}\left[w-W\right],\end{aligned} \tag{3.154}$$

where Eq.(3.149) has been used to obtain the last line.

### 3.3.3 Explicit evaluation of the entanglement entropy and energy for a tractable model in some stationary spacetimes

With the help of the formulas derived in the previous two subsections, we now calculate entanglement entropy and energy explicitly to construct entanglement thermodynamics for a tractable model in Minkowski, Schwarzschild and Reissner-Nordström spacetimes.

#### Model description

The basic idea of entanglement thermodynamics is to express the thermodynamic quantities for a black hole in terms of expectation values of quantum operators dependent on the spacetime division as in the statistical mechanics modeling of the thermodynamics for ordinary systems. Therefore we must specify how to divide spacetime into two regions and with respect to what kind of state the expectation values are taken.

According to the original idea of entanglement, it is clearly most natural to consider a dynamical spacetime describing black hole formation from a nearly flat spacetime in the past infinity, and divide the spacetime into the regions inside and outside the horizon. In this situation, if we start from the asymptotic Minkowski vacuum in the past, the entanglement entropy associated with the division of spacetime by the horizon acquires a clear physical meaning. However, this ideal modeling seems difficult.

Hence, we are obliged to consider a stationary spacetime, and take as the quantum state a stationary one. To be specific, we consider the vacuum with respect to a Killing time in a spherically symmetric static black hole described by a metric of the form

$$ds^2 = -N(\rho)^2 dt^2 + d\rho^2 + r(\rho)^2 \left(d\theta + \sin^2\theta d\psi^2\right). \tag{3.155}$$

The vacuum is well known as the Boulware state. As stated in subsections 3.2.1 and 3.2.2, the Boulware state has negative energy density and its contribution to gravitational mass diverges for a stationary black hole background. It has been shown in section 3.2 that the ground state of the brick wall model is the Boulware state and that the negative divergence in gravitational mass due to the Boulware state is canceled by a positive divergence due to thermal excitations above the Boulware state. Hence, we can safely state that the brick wall model is (the Boulware state) + (thermal excitations). Thermal characters of a black hole system can be explained by the latter: entropy and temperature of a black hole are modeled by entropy and temperature of the thermal excitations.

Our present purpose in this section is to explain the thermal characters of the black hole system in terms of entanglement. From the above arguments on the brick wall model, we expect that entanglement in the Boulware state does well for this purpose: entanglement entropy and entanglement temperature might explain entropy and temperature of a black hole; the negative divergence in gravitational mass due to the Boulware state might be canceled by entanglement energy, provided that the sum of the vacuum energy and entanglement energy contributes to gravitational mass.

Here a subtlety occurs: in, for example, a Schwarzschild spacetime, if we require that the quantum field is in the Boulware state on the whole extended Kruskal spacetime and take the bifurcation surface as the boundary surface, then the entanglement entropy vanishes because the state is expressed as the tensor product of the Boulware states in the regions I and II of *Figure* 3.2.

In order to avoid this, we restrict the spacetime into the region I, and replace the boundary by a timelike surface $\Sigma$ at a proper distance of the order of the cut-off length of the theory to the horizon (see *Figure* 3.2). This prescription corresponds to taking fluctuations of the horizon into consideration: quantum fluctuations of geometry near a horizon will prevent events closer to the horizon than about the Planck length from being seen on the outside. However, we should keep in mind that we have no definite criterion regarding the exact position of the boundary. To minimize this ambiguity, we will also investigate the influence of the variation of the boundary position.

As a matter content we consider a real scalar field described by

$$I = -\frac{1}{2}\int \sqrt{-g}dx^4[\partial^\mu\phi\partial_\mu\phi + m_\phi^2]. \tag{3.156}$$

The mass $m_\phi$ does not play an essential role since a typical length scale controlling the entanglement thermodynamics is much smaller than the Compton length of an usual field. Therefore we just set $m_\phi = 0$ in the following arguments.

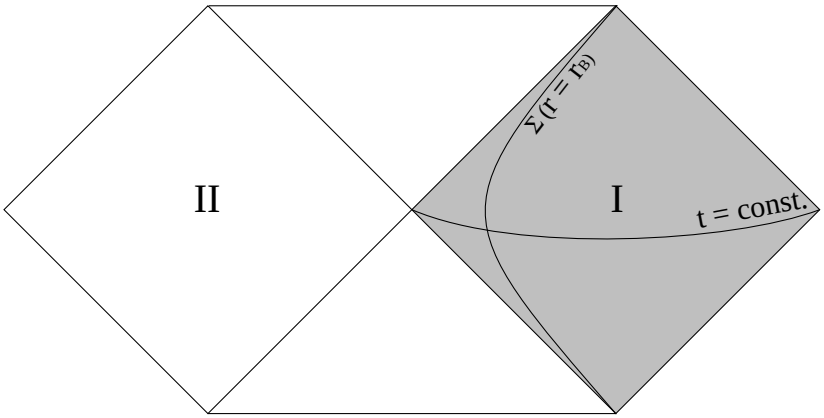


Figure 3.2: The Kruskal extension of the Schwarzschild spacetime. We consider only the region $I$ (the shaded region). As the boundary $\Sigma$ we take the hypersurface $r = r_B$.

**Discretized theory of a scalar field**

Consider a massless real scalar field described by the action

$$I = -\frac{1}{2}\int dx^4 \sqrt{-g}\partial^\mu\phi\partial_\mu\phi, \tag{3.157}$$

where the background geometry is fixed to be a spherically symmetric static spacetime with the metric (3.155). For this system we calculate entanglement entropy and entanglement energy to construct entanglement thermodynamics using the methods developed in the previous two subsections. Those methods are both based on a discrete system $\{q^A\}$ $(A = 1, 2, \cdots, n_{tot})$ described by a Hamiltonian of the form (3.115). For this discrete system it is easy to divide the whole Hilbert space $\mathcal{F}$ into the form (3.109): $\mathcal{F}$ is defined as a Fock space constructed from $\{q^A\}$ $(A = 1, 2, \cdots, n_{tot})$; $\mathcal{F}_1$ is defined as a Fock space constructed from $\{q^a\}$ $(a = 1, 2, \cdots, n_B)$; $\mathcal{F}_2$ is defined as a Fock space constructed from $\{q^\alpha\}$ $(\alpha = n_B + 1, \cdots, n_{tot})$. In order to apply this scheme to our problem we have to construct a discretized theory of the scalar field whose Hamiltonian is of the form (3.115).

First we expand the field $\phi$ in terms of the spherical harmonics as

$$\phi(\rho, \theta, \psi) = \sum_{l,m} \frac{N^{1/2}}{r}\phi_{lm}(\rho)Z_{lm}(\theta, \psi), \tag{3.158}$$

where $Z_{lm} = \sqrt{2}\Re Y_{lm}$ for $m > 0$, $\sqrt{2}\Im Y_{lm}$ for $m < 0$, and $Z_{l0} = Y_{l0}$. Then the Hamiltonian corresponding to the Killing time for the action (3.157) is decomposed into a direct sum of contributions from each harmonics component $H_{lm}$ as

$$H = \sum_{lm} H_{lm} \quad . \tag{3.159}$$

Here $H_{lm}$ is given by

$$H_{lm} = \frac{1}{2}\int d\rho \left[\pi_{lm}^2 + Nr^2\left\{\frac{\partial}{\partial\rho}\left(\frac{N^{1/2}}{r}\phi_{lm}\right)\right\}^2 + l(l+1)\left(\frac{N\phi_{lm}}{r}\right)^2\right] \quad , \quad (3.160)$$

where $\pi_{lm}(\rho)$ is a momentum conjugate to $\phi_{lm}(\rho)$.

Note that for any $(l,m)$ the Hamiltonian (3.160) of the subsystem has the form

$$H_{lm} = \frac{1}{2}\int d\rho \pi_{lm}^2(\rho) + \frac{1}{2}\int d\rho d\rho' \phi_{lm}(\rho) V^{(l,m)}(\rho,\rho')\phi_{lm}(\rho'), \quad (3.161)$$

where the following algebra of Poisson brackets is understood:

$$\begin{aligned} \{\phi_{lm}(\rho), \pi_{l'm'}(\rho')\} &= \delta(\rho-\rho'), \\ \{\phi_{lm}(\rho), \phi_{l'm'}(\rho')\} &= 0, \\ \{\pi_{lm}(\rho), \pi_{l'm'}(\rho')\} &= 0. \end{aligned} \quad (3.162)$$

Each subsystem described by the Hamiltonian (3.161) can be discretized by the following procedure:

$$\begin{aligned} \rho &\rightarrow (A-1/2)a \quad , \\ \delta(\rho-\rho') &\rightarrow \delta_{AB}/a \quad , \end{aligned} \quad (3.163)$$

where $A, B = 1, 2, \cdots$ and $a$ is a cut-off length. The corresponding Hamiltonian of the discretized system is of the form (3.115) with

$$\begin{aligned} \phi_{lm}(\rho) &\rightarrow q^A \quad , \\ \pi_{lm}(\rho) &\rightarrow p_A/a \quad , \\ V^{(l,m)}(\rho,\rho') &\rightarrow V_{AB}/a^2 \quad . \end{aligned} \quad (3.164)$$

In this way we obtain a discretized system with the total Hamiltonian (3.115) with the matrix $V$ given by the direct sum

$$V = \bigoplus_{l,m} V^{(l,m)}, \quad (3.165)$$

where $V^{(l,m)}$ is independent of $m$ and is explicitly expressed as

$$\begin{aligned} V_{AB}^{(l,m)}\phi_{lm}^A\phi_{lm}^B &= a\sum_{A=1}^{\infty}\left[N_{A+1/2}\left(\frac{x_{A+1/2}}{a}\right)^2\left(\frac{N_{A+1}^{1/2}}{x_{A+1}}\phi_{lm}^{A+1} - \frac{N_A^{1/2}}{x_A}\phi_{lm}^A\right)^2\right. \\ &\left. + \frac{l(l+1)}{r_0^2}\left(\frac{N_A\phi_{lm}^A}{x_A}\right)^2\right] \quad . \end{aligned} \quad (3.166)$$

Here

$$\begin{aligned} x_A &= r(\rho=(A-1/2)a)/r_0 \quad , \\ x_{A+1/2} &= r(\rho=Aa)/r_0 \quad , \\ N_A &= N(\rho=(A-1/2)a) \quad , \\ N_{A+1/2} &= N(\rho=Aa) \quad , \\ \phi_{lm}^A &= \phi_{lm}(\rho=(A-1/2)a) \quad . \end{aligned} \quad (3.167)$$

In the matrix representation $V^{(l,m)}$ is given by the $n_{tot} \times n_{tot}$ matrix

$$\begin{aligned}
\left(V_{AB}^{(l,m)}\right) &= \frac{2a}{r_0^2}\begin{pmatrix} \Sigma_1^{(l)} & \Delta_1 & & & & \\ \Delta_1 & \Sigma_2^{(l)} & \Delta_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \Delta_{A-1} & \Sigma_A^{(l)} & \Delta_A & \\ & & & \ddots & \ddots & \ddots \end{pmatrix} , \\
\Sigma_A^{(l)} &= \frac{1}{2}(r_0/a)^2 N_A x_A^{-2}\left[N_{A-1/2}x_{A-1/2}^2 + N_{A+1/2}x_{A+1/2}^2\right] \\
&\quad + \frac{1}{2}l(l+1)N_A^2 x_A^{-2} \ , \\
\Delta_A &= -\frac{1}{2}(r_0/a)^2 N_A^{1/2} N_{A+1/2} N_{A+1}^{1/2} x_A^{-1} x_{A+1/2}^2 x_{A+1}^{-1} \ , \qquad (3.168)
\end{aligned}$$

where we have imposed the boundary condition $\phi_{lm}^{n_{tot}+1} = 0$. In these expressions $r_0$ is an arbitrary constant, which we set to be area radius of a horizon [7] for convenience in the following arguments.

### Spatial division

We divide the total system by a stationary hypersurface $\Sigma$ defined by $r = r_B$: we split the system $\{\phi_{lm}^A\}$ $(A = 1, \cdots, n_{tot})$ into the two subsystems, $\{\phi_{lm}^a\}$ $(a = 1, \cdots, n_B)$ and $\{\phi_{lm}^\alpha\}$ $(\alpha = n_B + 1, \cdots, n_{tot})$; the area radius $r_B$ of the boundary is given by $r_B = r(\rho = n_B a)$.

Since this division preserves spherical symmetry of the system, we can still apply the expansion by harmonics. Thus, entanglement quantities of the system are calculated as sums of all contributions from those subsystems, each of which is specified by $(l, m)$ and described by the matrix $V^{(l,m)}$.

### Convergence of the summation

Since $V^{(l,m)}$ is independent of $m = (-l, -l+1, \cdots, l-1, l)$, the entanglement entropy and energy are given by

$$\begin{aligned}
S_{ent} &= \sum_{l=0}^{\infty}(2l+1)S_{ent}^{(l)}, \\
E_{ent} &= \sum_{l=0}^{\infty}(2l+1)E_{ent}^{(l)}, \qquad (3.169)
\end{aligned}$$

where $S_{ent}^{(l)}$ and $E_{ent}^{(l)}$ are entanglement entropy and energy of the subsystem specified by $(l, m)$ and independent of $m$.

From Eq.(3.168), one can easily show that

$$\begin{aligned}
S_{ent}^{(l)} &\sim O\left((la/r_0)^{-4}\ln(la/r_0)\right) \quad \text{as} \quad la/r_0 \to \infty, \\
E_{ent}^{(l)} &\sim O\left((la/r_0)^{-3}\right) \quad \text{as} \quad la/r_0 \to \infty. \qquad (3.170)
\end{aligned}$$

Thus the infinite sums Eq.(3.169) actually converge so that we can safely truncate them at some appropriate $l$, depending on the accuracy we require and the ratio $r_0/a$ we set.

[7] For the model in Minkowski spacetime we take $r_0 = a$.

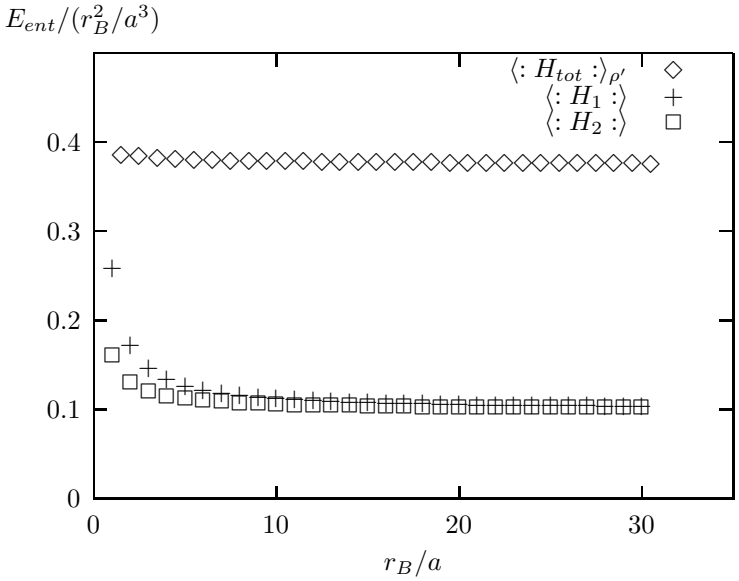


Figure 3.3: The numerical evaluations of entanglement energy in Minkowski spacetime. $E_{ent}/(r_B^2/a^3)$ is shown as functions of $r_B/a$, where $E_{ent}$ denotes $\langle : H_1 : \rangle$, $\langle : H_2 : \rangle$ and $\langle : H_{tot} : \rangle_{\rho'}$, respectively, and $r_B \equiv n_B a$. We have taken $n_{tot} = 60$ for $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$, and $n_{tot} = 200$ for $\langle : H_{tot} : \rangle_{\rho'}$.

**Numerical calculations in Minkowski spacetime**

Before analyzing our model in black hole geometries, it is instructive to calculate entanglement quantities in the simple models in Minkowski spacetime introduced by Bombelli et.al. [37] and Srednicki [38]. They calculated entanglement entropy for divisions of a flat hypersurface in Minkowski spacetime by $\mathbf{R}^2$ and $S^2$, respectively. Their results are summarized as Eq. (3.130).

Here, let us calculate entanglement energy in the model of Srednicki (a division by $S^2$) [8]. The discretized theory of a scalar field corresponding to this model can be easily obtained by the above procedure by setting $r_0 = a$, $r(\rho) = \rho$ and $N(\rho) = 1$. The discretized system $\{\phi^A_{lm}\}$ $(A = 1, \cdots, n_{tot})$ is described by Hamiltonian of the form (3.115) with the potential given by (3.168).

*Figure* 3.3 shows the result of numerical calculations of $\langle : H_1 : \rangle$, $\langle : H_2 : \rangle$ and $\langle : H_{tot} : \rangle_{\rho'}$ in this model.

From this figure we see that $E_{ent}$ is almost proportional to $r_B^2/a^3$:

$$\begin{aligned} \langle : H_1 : \rangle &\simeq \langle : H_2 : \rangle \simeq 0.1 \frac{r_B^2}{a^3}, \\ \langle : H_{tot} : \rangle_{\rho'} &\simeq 0.4 \frac{r_B^2}{a^3}. \end{aligned} \tag{3.171}$$

**Numerical calculations in Schwarzschild spacetime**

In Schwarzschild spacetime the metric is given by

$$ds^2 = -\left(1 - \frac{r_0}{r}\right) dt^2 + \left(1 - \frac{r_0}{r}\right)^{-1} dr^2 + r^2 \left(d\theta + \sin^2\theta d\psi^2\right), \tag{3.172}$$

[8] Entanglement energy in the model of Bombelli et.al (a division by $\mathbf{R}^2$) is analyzed in Appendix A.5.

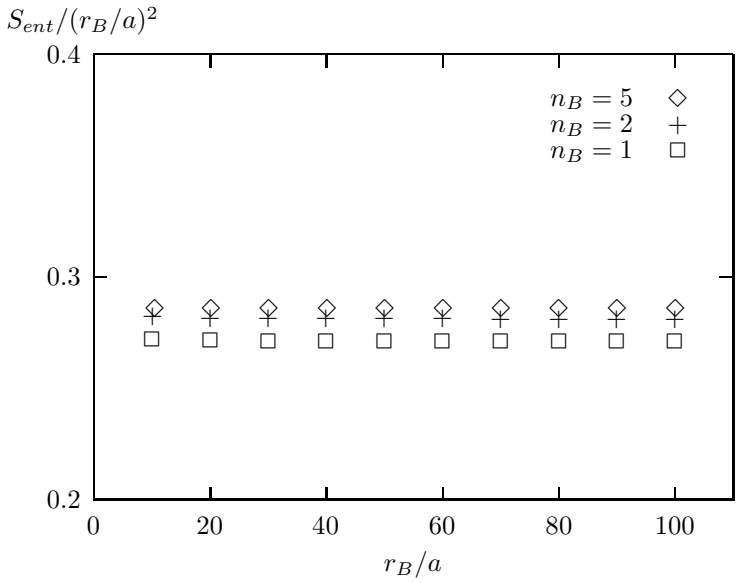


Figure 3.4: The numerical evaluations for $S_{ent}$ of the discretized theory of the scalar field in Schwarzschild spacetime. $S_{ent}/(r_B/a)^2$ for $n_B = 1, 2, 5$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

where $r_0$ is the area radius of the horizon. As the radial coordinate $\rho$ we take the proper distance from the horizon:

$$\rho = \frac{r_0}{2}\left[\sqrt{y^2-1} + \ln\left(y + \sqrt{y^2-1}\right)\right] \quad , \tag{3.173}$$

where the variable $y$ is defined by $y = 2r/r_0 - 1$.

Using formulas given in subsection 3.3.1, we have evaluated $S_{ent}$ numerically. In this calculation the outer numerical boundary is set at $n_{tot} = 100$. The summation with respect to $l$ in Eq.(3.169) is taken up to $l = [10r_0/a]$ ([ ] is the Gauss symbol). From the above asymptotic behavior of $S_{ent}^{(l)}$, this guarantees the accuracy of 10%.

The result is shown in *Figure* 3.4. From this figure we see that $S_{ent}$ is proportional to $(r_B/a)^2$ if we change $r_0$ with $n_B$ fixed, and its coefficient has only a weak dependence on $n_B$. Thus, we get

$$S_{ent} \simeq 0.3\left(\frac{r_B}{a}\right)^2. \tag{3.174}$$

This result is essentially the same as the previous result (3.130) for models in Minkowski spacetime including the numerical coefficient. This can be understood in the following way.

Let us make a coordinate change from $r$ to $x$ defined by

$$\frac{r}{r_0} = \frac{(x+1)^2}{4x},$$

or

$$x = \frac{2r}{r_0} - 1 + \sqrt{\left(\frac{2r}{r_0} - 1\right)^2 - 1}.$$

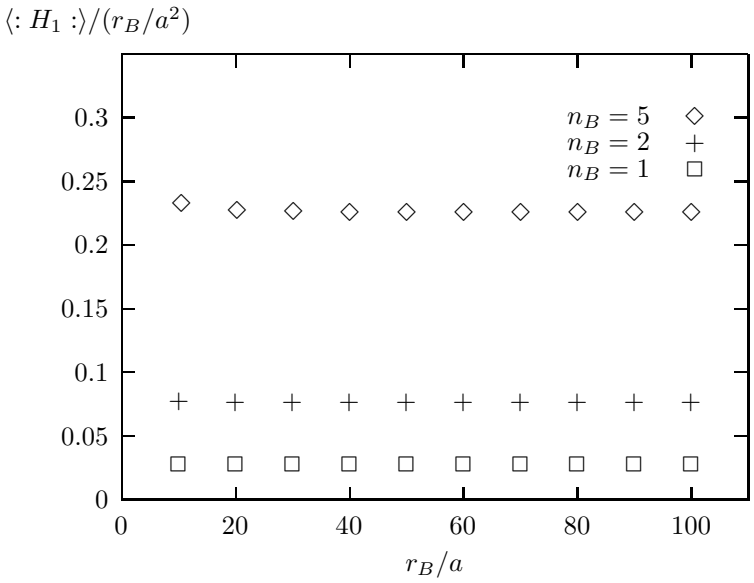


Figure 3.5: The numerical evaluations of $\langle : H_1 : \rangle$ for the discretized theory of the scalar field in Schwarzschild spacetime. $\langle : H_1 : \rangle / (r_B/a^2)$ for $n_B = 1, 2, 5$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

Then the metric (3.172) is rewritten as

$$ds^2 = -\left(\frac{x-1}{x+1}\right)^2 dt^2 + r_0^2 \left(\frac{1+x}{2x}\right)^4 (dx^2 + x^2 d\Omega^2).$$

Note that $r = 0$, $r_0$ and $\infty$ correspond to $x = 0$, 1 and $\infty$, respectively. It is easy to see that the Hamiltonian is given in this coordinate system as

$$\begin{aligned} H &= \sum_{lm} H_{lm}, \\ H_{lm} &= \int d\xi \frac{64x^4(x-1)}{(x+1)^7} \left[ \frac{1}{2} P_{lm}^2 \right. \\ &\quad \left. + \frac{1}{2} \left( \frac{(x+1)^2}{4x} \right)^4 \left\{ (\partial_\xi \varphi_{lm})^2 + \frac{l(l+1)}{\xi^2} \varphi_{lm}^2 \right\} \right], \end{aligned} \tag{3.175}$$

where $\xi := r_0 x$, and $P_{lm}$ and $\varphi_{lm}$ are expressed as $P_{lm} := \frac{r_0}{64} \frac{(x+1)^7}{x^4(x-1)} \dot{\phi}_{lm}$ and $\varphi_{lm} := r_0 \phi_{lm}$ in terms of $\phi_{lm}$. Here note that the vacuum state is only weakly dependent on the prefactor $\frac{64x^4(x-1)}{(x+1)^7}$ in Eq.(3.175). If we neglect this prefactor, the vacuum state is determined by the Hamiltonian which coincides with that for the flat spacetime at $x = 1$. On the other hand, $S_{ent}$ depends on the modes in a thin layer around the boundary $\Sigma$, whose typical thickness is a few times of $a \simeq l_{\rm Pl}$. Therefore, when $\Sigma$ is near the horizon, the value of $S_{ent}$ should be well approximated by the flat spacetime value.

Next, with the helps of formulas developed in subsection 3.3.2, we have numerically evaluated $\langle : H_1 : \rangle$, $\langle : H_2 : \rangle$ and $\langle : H_{tot} : \rangle_{\rho'}$. Now we have taken the numerical outer boundary at $n_{tot} = 100$. The truncation in the $l$-summation is the same as

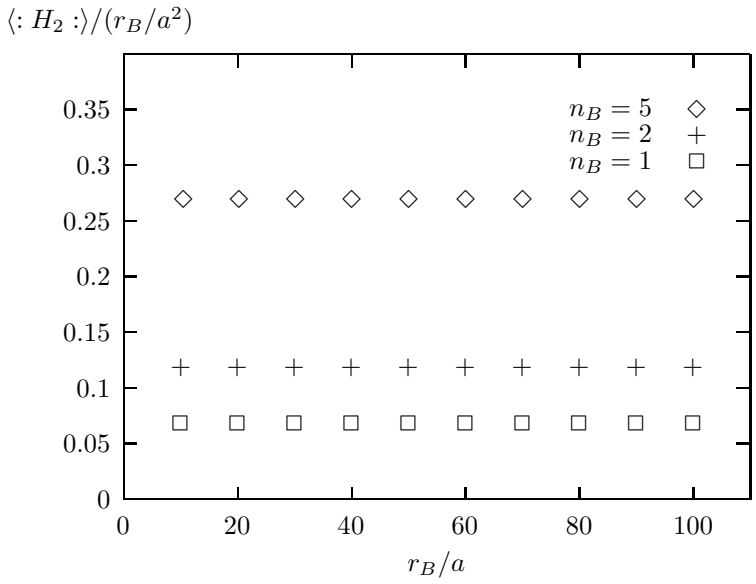


Figure 3.6: The numerical evaluations of $\langle : H_2 : \rangle$ for the discretized theory of the scalar field in Schwarzschild spacetime. $\langle : H_2 : \rangle / (r_B/a^2)$ for $n_B = 1, 2, 5$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

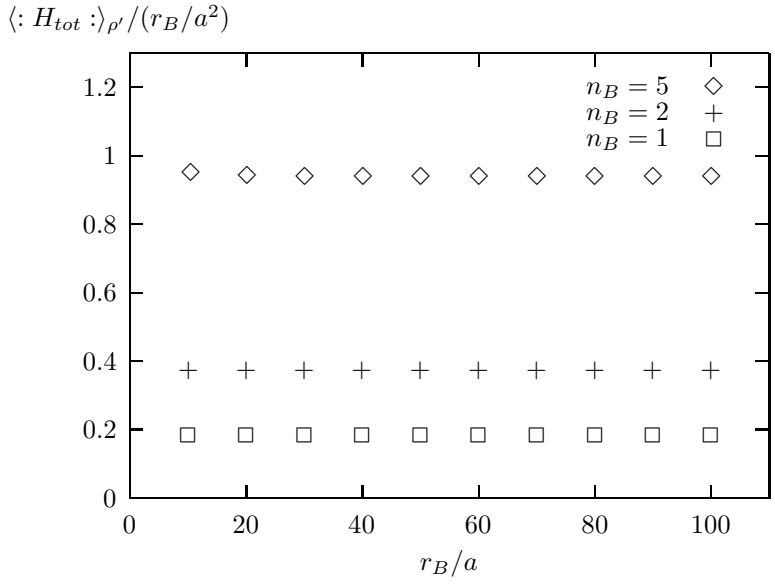


Figure 3.7: The numerical evaluations of $\langle : H_{tot} : \rangle_{\rho'}$ for the discretized theory of the scalar field in Schwarzschild spacetime. $\langle : H_{tot} : \rangle_{\rho'} / (r_B/a^2)$ for $n_B = 1, 2, 5$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

for $S_{ent}$ (up to $l = [10r_0/a]$), which implies that the accuracy is about 10% from the above asymptotic estimate for $E^{(l)}_{ent}$.

The results of numerical calculations are shown in *Figure* 3.5, *Figure* 3.6 and *Figure* 3.7. In these figures, $E_{ent}/(r_B/a^2)$ is plotted as a function of $r_B/a$ for $n_B = 1, 2, 5$. All of these figures show that $E_{ent}$ is proportional to $r_B/a^2$:

$$\begin{aligned} \langle : H_1 : \rangle &\simeq 0.05(n_B - 1/2)r_B/a^2, \\ \langle : H_2 : \rangle &\simeq 0.05(n_B + 1/2)r_B/a^2, \\ \langle : H_{tot} : \rangle_{\rho'} &\simeq 0.2 n_B r_B/a^2. \end{aligned} \tag{3.176}$$

From these equations we immediately see that the values of $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$ coincide except for a tiny difference independent of $n_B$. This difference is understood by the gravitational red-shift: $\langle : H_1 : \rangle$ comes from the modes just inside $\Sigma$ while $\langle : H_2 : \rangle$ originates from the modes just outside $\Sigma$. In the present numerical calculations, it means that $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$ are determined by the modes at $\rho = (n_B - 1/2)a$ and $\rho = (n_B + 1/2)a$, respectively(see Eq.(3.168)). Hence, taking account of the fact that the contribution of each mode to the entanglement energy is proportional to the red-shift factor at its location, the ratio of $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$ should be approximately given by

$$\begin{aligned} \langle : H_1 : \rangle : \langle : H_2 : \rangle &\sim N^{1/2}(\rho = (n_B - 1/2)a) : N^{1/2}(\rho = (n_B + 1/2)a) \\ &\sim (n_B - 1/2) : (n_B + 1/2). \end{aligned} \tag{3.177}$$

This is consistent with the above numerical result.

This argument is also supported by the numerical result for the flat spacetime model shown in *Figure* 3.3. In this figure the values of $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$ for a massless scalar field in the Minkowski spacetime with $\Sigma = B \times \mathbf{R} = S^2 \times \mathbf{R}$ are plotted. In this case there is no gravitational red-shift effect, so we expect that $\langle : H_1 : \rangle \simeq \langle : H_2 : \rangle$ as confirmed by the numerical calculation.

**Numerical calculations in Reissner-Nordström spacetime**

In Reissner-Nordström spacetime the metric is given by

$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right) dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\psi^2\right), \tag{3.178}$$

where $M$ and $Q$ are the mass and the charge of the black-hole. The area radius of the outer horizon $r_0$ is

$$r_0 = M + \sqrt{M^2 - Q^2}. \tag{3.179}$$

As the radial coordinate $\rho$ we take the proper distance from the outer horizon

$$\rho = \sqrt{(M^2 - Q^2)(y^2 - 1)} + M \ln\left(y + \sqrt{y^2 - 1}\right), \tag{3.180}$$

where the variable $y$ is defined by

$$y = \frac{r - M}{\sqrt{M^2 - Q^2}} \quad . \tag{3.181}$$

Using formulas given in subsection 3.3.1, we have evaluated $S_{ent}$ numerically. In this calculation the outer numerical boundary is set at $n_{tot} = 100$, and the boundary of the spatial division is fixed at $n_B = 1$. The summation in $l$ in Eq.(3.169) is taken up to $l = [10r_0/a]$ ([ ] is the Gauss symbol). From the above asymptotic behavior of $S^{(l)}_{ent}$, this guarantees the accuracy of 10%.

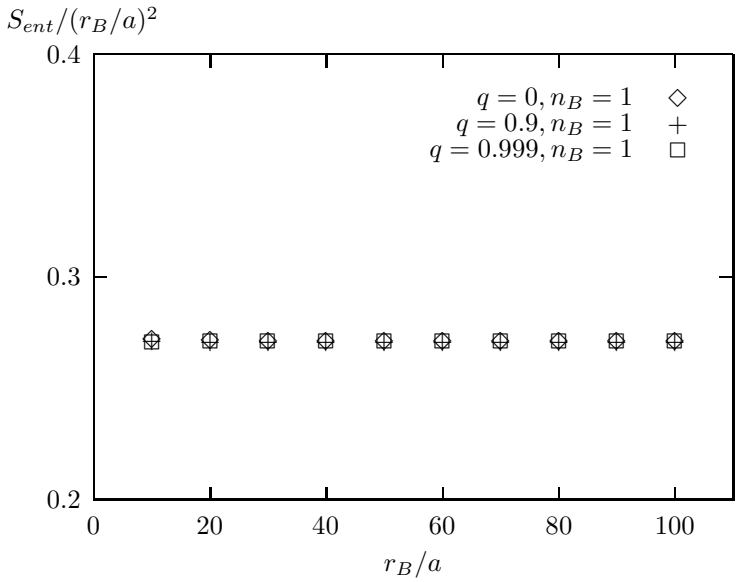


Figure 3.8: The numerical evaluations of $S_{ent}$ for the discretized theory of the scalar field in Reissner-Nordström spacetime. $S_{ent}/(r_B/a)^2$ for $q = 0, 0.9, 0.999$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

The result is shown in *Figure* 3.8. From this figure we see that $S_{ent}$ is proportional to $(r_B/a)^2$ if we change $r_0$ with $q \equiv Q/M$ fixed, and its coefficient has only a weak dependence on $q$. Thus, we get

$$S_{ent} \simeq 0.3 \left(\frac{r_B}{a}\right)^2. \tag{3.182}$$

This result is essentially the same as the results in Minkowski and Schwarzschild spacetime.

Next, by using formulas in subsection 3.3.2, we have numerically evaluated $\langle: H_1 :\rangle$, $\langle: H_2 :\rangle$ and $\langle: H_{tot} :\rangle_{\rho'}$. We have taken the numerical outer boundary at $n_{tot} = 100$ and the boundary of the spatial division at $n_B = 1$. The truncation in the $l$-summation is the same as for $S_{ent}$ (up to $l = [10r_0/a]$), which implies that the accuracy is about 10% from the above asymptotic estimate for $E^{(l)}_{ent}$.

The results of numerical calculations are shown in *Figure* 3.9, *Figure* 3.10 and *Figure* 3.11. In these figures, $E_{ent}/(c(q)r_B/a^2)$ is plotted as a function of $r_B/a$ for $n_B = 1$ and $q = 0, 0.9, 0.999$, where $c(q)$ is defined by

$$c(q) = \frac{2\sqrt{1-q^2}}{1+\sqrt{1-q^2}}. \tag{3.183}$$

All of these figures show that $E_{ent}$ is proportional to $c(q)r_B/a^2$. The result is summarized as

$$\begin{aligned} \langle: H_1 :\rangle &\simeq 0.025 c(q) r_B/a^2, \\ \langle: H_2 :\rangle &\simeq 0.075 c(q) r_B/a^2, \\ \langle: H_{tot} :\rangle_{\rho'} &\simeq 0.2 c(q) r_B/a^2. \end{aligned} \tag{3.184}$$

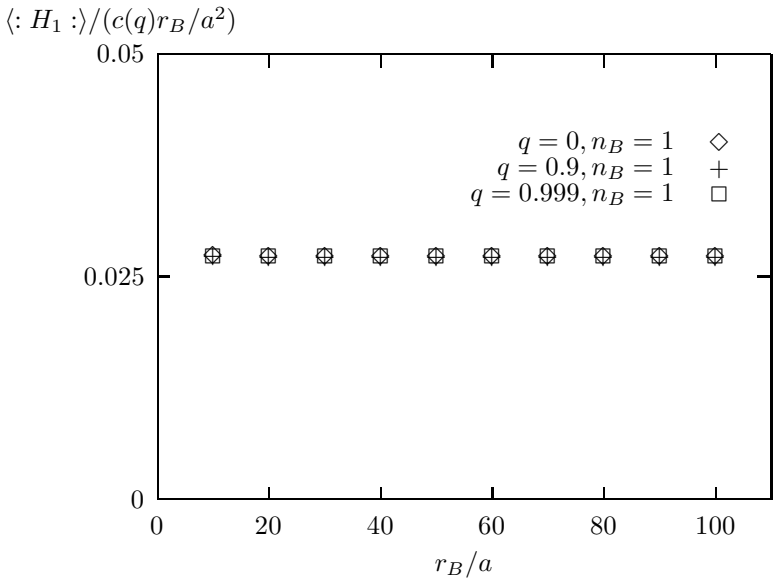


Figure 3.9: The numerical evaluations of $\langle : H_1 : \rangle$ for the discretized theory of the scalar field in Reissner-Nordström spacetime. $\langle : H_1 : \rangle / (c(q) r_B / a^2)$ for $q = 0, 0.9, 0.999$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10 r_0/a$.

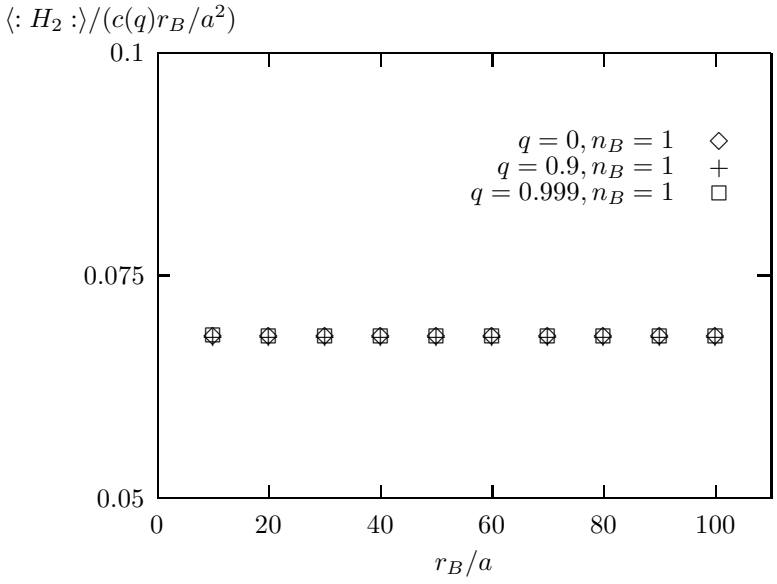


Figure 3.10: The numerical evaluations of $\langle : H_2 : \rangle$ for the discretized theory of the scalar field in Reissner-Nordström spacetime. $\langle : H_2 : \rangle / (c(q) r_B / a^2)$ for $q = 0, 0.9, 0.999$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10 r_0/a$.

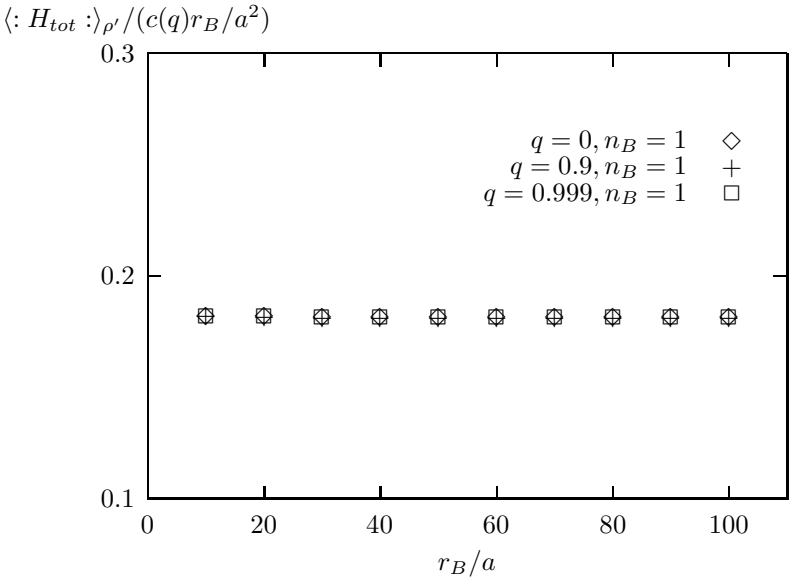


Figure 3.11: The numerical evaluations of $\langle : H_{tot} : \rangle_{\rho'}$ for the discretized theory of the scalar field in Reissner-Nordström spacetime. $\langle : H_{tot} : \rangle_{\rho'}/(c(q)r_B/a^2)$ for $q = 0, 0.9, 0.999$ is shown as functions of $r_B/a$, where $r_B \equiv r(\rho = n_B a)$. We have taken $n_{tot} = 100$ and performed the summation over $l$ up to $10r_0/a$.

### 3.3.4 Comparison: entanglement thermodynamics and black-hole thermodynamics

**Entanglement thermodynamics in Minkowski spacetime**

We have introduced four possible definitions of entanglement energy. By combining each of them with the entanglement entropy $S_{ent}$, we obtain entanglement thermodynamics.

For this purpose, we consider an infinitesimal process in which the way of the division of the Hilbert space $\mathcal{H}$ into $\mathcal{H}_1$ and $\mathcal{H}_2$ is changed smoothly, with the 'initial' state $\rho_0$ being fixed. (See subsection 3.3.1.) Let $\delta S_{ent}$ and $\delta E_{ent}$ be the resultant infinitesimal changes in the entanglement entropy and in the entanglement energy, respectively. We are dealing with a 1-parameter family of the infinitesimal changes for the entanglement. The parameter is chosen to be the area radius of the boundary sphere. Thus, the construction of the thermodynamics means to use the first law of entanglement thermodynamics (3.108) to determine $T_{ent}$, which we call entanglement temperature. Combining (3.130) and (3.171) with Eq.(3.108), we thus get [9]

$$k_B T_{ent} = \frac{\mathcal{N}_E}{\mathcal{N}_S} \cdot \frac{\hbar c}{a}, \tag{3.185}$$

where $\mathcal{N}_E$ is a numerical factor in Eq. (3.171). Note that entanglement temperatures $T_{ent}$ obtained from the four definitions of the entanglement energy coincides up to numerical factors of order unity.

Let us interpret entanglement thermodynamics given by (3.130), (3.171) and

[9] In this subsection, we recover $\hbar$ and $c$.

(3.185). It is helpful to introduce the quantities

$$\begin{aligned} n_{ent} &\equiv \left(\frac{r_B}{a}\right)^2, \\ e_{ent} &\equiv \frac{\hbar c}{a}. \end{aligned} \tag{3.186}$$

Here $n_{ent}$ is regarded as an effective number of degrees of freedom of matter on the boundary 2-surface $B$, and $e_{ent}$ is a typical energy scale of each degree of freedom on $B$.

From Eqs. (3.130), (3.171) and (3.185), we find that

$$\begin{aligned} S_{ent} &\sim k_B n_{ent}, \\ E_{ent} &\sim e_{ent} n_{ent}, \\ k_B T_{ent} &\sim e_{ent}. \end{aligned} \tag{3.187}$$

Therefore our results can be interpreted as follows[10] : The entanglement entropy is a measure for the number of microscopic degrees of freedom on the boundary $B$; the entanglement energy is a measure for the total energy carried by all of the degrees of freedom on $B$; entanglement temperature is measure for the energy carried by each degree of freedom on $B$.

**Discrepancy between entanglement thermodynamics in Minkowski spacetime and black-hole thermodynamics**

Now we compare these results with the case of black holes. For that purpose we express the black-hole thermodynamics in the same form as in the previous subsection.

Let us introduce the quantities

$$\begin{aligned} n_{BH} &\equiv \left(\frac{r_0}{l_{pl}}\right)^2, \\ e_{BH} &\equiv \frac{\hbar c}{l_{pl}}, \end{aligned} \tag{3.188}$$

where $r_0$ is area radius of an event horizon of a black hole. We can interpret that $n_{BH}$ corresponds to the effective number of degrees of freedom on the event horizon and $e_{BH}$ is a typical energy scale for each degree of freedom of matter on the horizon.

The black-hole thermodynamics can be recast in terms of these quantities as

$$\begin{aligned} S_{BH} &\sim k_B n_{BH}, \\ E_{BH} &\sim \gamma_{BH} e_{BH} n_{BH}, \\ k_B T_{BH} &\sim \gamma_{BH} e_{BH}, \end{aligned} \tag{3.189}$$

where $\gamma_{BH} \equiv l_{pl}/r_0$. The factor $\gamma_{BH}$ can be understood as a magnification of energy due to an addition of gravitational energy or a red-shift factor of temperature since $\sqrt{-g_{tt}} \sim l_{pl}/r_0$ at $r \sim r_0 + l_{pl}^2/r_0$, which corresponds to a stationary observer at the proper distance $l_{pl}$ away from the horizon. Thus the following interpretation is possible[11]: The black-hole entropy is a measure for the number of the microscopic degrees of freedom on the event horizon; the black-hole energy is a measure at infinity for the total energy carried by all of the degrees of freedom on the event

[10] It is safer, however, to regard such an interpretation just as a convenient way of representing our results. (See the next section for another interpretation of entanglement entropy. This note in particular applies to the case of the black-hole thermodynamics (see Eq.(3.189)).

[11] See the footnote after Eq.(3.187).

horizon; the black-hole temperature is a measure at infinity for the energy carried by each degree of freedom.

Now we compare the two types of thermodynamics characterized by Eq.(3.187) and Eq.(3.189), respectively. Both of them allow the interpretation that they describe the behavior of the effective microscopic degrees of freedom on the boundary $B$, or on the horizon. Because of the factor $\gamma_{BH}$, however, they are hardly understood in a unified picture. This strongly suggests that an inclusion of gravitational effects is necessary for agreement between them.

The discrepancy is highlighted in the context of the third law of thermodynamics. Both types of thermodynamics fail to follow Planck's version of the third law [12], but in quite different manners.

From Eq.(3.185), we see that $T_{ent}$ remains constant if $\frac{\mathcal{N}_E}{\mathcal{N}_S}$ is assumed to be constant [13]. On the other hand, Eq.(3.130) shows that $S_{ent}$ tends to zero as $A \to 0$, where $A = 4\pi r_B^2$ is area of the boundary. Therefore we obtain the following $A$-dependence:

$$\begin{aligned} S_{ent} &\propto A, \\ E_{ent} &\propto A, \\ k_B T_{ent} &\propto A^0. \end{aligned} \tag{3.190}$$

The system behaves as though it is kept in touch with a thermal bath with temperature $T_{ent}$.

In contrast, for the black-hole thermodynamics, Eq.(1.5) and Eq.(1.6) along with Eq.(1.7) give the behavior (note that $A \propto M_{BH}^2$)

$$\begin{aligned} S_{BH} &\propto A, \\ E_{BH} &\propto \sqrt{A}, \\ k_B T_{BH} &\propto 1/\sqrt{A}, \end{aligned} \tag{3.191}$$

where $A = 4\pi r_0^2$ is area of the horizon. Thus we see that $S_{BH} \to \infty$ as $T_{BH} \to 0$.

The discrepancy between Eq.(3.190) and Eq.(3.191) is quite impressive. On one hand, a well-known behavior (3.191) comes from the fundamental properties of the black-hole physics. On the other hand, the behavior characterized by Eq.(3.190) is also an universal one in any model of the entanglement: the zero-point energy of the system has been subtracted in the definitions of entanglement energy, thus only the degrees of freedom on the boundary $B$ contributes to $E_{ent}$, yielding the behavior $E_{ent} \propto A$. The definitions of $E_{ent}$ proposed here look quite reasonable though other definitions may be possible. The result $E_{ent} \propto A$ also looks natural, being compatible with the concept of 'entanglement'. At the same time, $S_{ent}$ also behaves universally as $S_{ent} \propto A$, which has been the original motivation for investigating the relation between $S_{BH}$ and $S_{ent}$ [37, 38, 39].

### Restoration of the agreement by inclusion of gravity

Let us discuss a possible restoration of the agreement between entanglement thermodynamics and black-hole thermodynamics by considering gravitational effects.

Although the behavior (3.187) of entanglement thermodynamics was derived by considering models in flat spacetime, it seems very reasonable that we regard the quantities $S_{ent}$, $E_{ent}$ and $T_{ent}$ as those in a black-hole background measured by a

[12] Here we mention that the black hole system does satisfy the third law in the sense of Nernst [8, 14].

[13] Here we are regarding the cut-off scale $a$ as the fundamental constant of the theory, not to be varied.

stationary observer located at the proper distance $a$ away from the horizon [14]. Since $S_{BH}$, $E_{BH}$ and $T_{BH}$ in (3.189) are quantities measured at infinity, it is behavior of $S_{ent}$, $E_{ent}$ and $T_{ent}$ at infinity that we have to compare with (3.189).

$S_{ent}$ at infinity probably has the same behavior as that measured by the observer near the horizon since a number of degrees of freedom seems independent of an observer's view-point. That is consistent with the fact that the entanglement entropy on Schwarzschild background has the same behavior $S_{ent} \sim k_B n_{ent}$ [39, 46]. On the other hand it seems natural to add the gravitational energy to the entanglement energy by replacing $E_{ent}$ with $\sqrt{-g_{tt}}E_{ent}$. Then entanglement temperature is determined by use of the first law (3.108). Thus the inclusion of gravity may alter the behavior (3.187) to

$$\begin{aligned} S_{ent} &\sim k_B n_{ent} \quad , \\ E_{ent} &\sim \gamma_{ent} e_{ent} n_{ent} \quad , \\ k_B T_{ent} &\sim \gamma_{ent} e_{ent} \quad , \end{aligned} \tag{3.192}$$

where $\gamma_{ent} \equiv a/r_0$. The factor $\gamma_{ent}$ represents the gravitational magnification of the entanglement energy due to the addition of gravitational energy since on the corresponding Schwarzschild background $\sqrt{-g_{tt}} \sim a/r_0$ at $r \sim r_0 + a^2/r_0$, which corresponds to a stationary observer at the proper distance $a$ away from the horizon (see the argument below (3.189)). Here $r_0$ is the area radius of the horizon.

The revised behavior (3.192) shows a complete agreement with (3.189), provided that $r_B \simeq r_0$ and $a \simeq l_{pl}$. Note that the last equality in (3.192) is consistent with an interpretation that the entanglement temperature is red-shifted by the factor $\gamma_{ent}$. Thus the inclusion of gravitational effects restores the agreement between the entanglement thermodynamics and the black-hole thermodynamics at least qualitatively.

**Entanglement thermodynamics in Schwarzschild spacetime**

From the numerical results in subsection 3.3.3, entanglement entropy and entanglement energy in Schwarzschild spacetime are expressed as

$$\begin{aligned} S_{ent} &\simeq k_B \mathcal{N}_S \left(\frac{r_B}{a}\right)^2 , \\ E_{ent} &\simeq \hbar c \mathcal{N}_E \frac{r_B}{a^2}, \end{aligned} \tag{3.193}$$

for all definitions of $E_{ent}$, where $r_B$ is the area radius of the boundary defined by $r_B = r(\rho = n_B a)$. Here $\mathcal{N}_S$ and $\mathcal{N}_E$ are numerical factors of order 1: $\mathcal{N}_S = 0.3$; $\mathcal{N}_E = 0.05(n_B - 1/2)$, $0.05(n_B + 1/2)$, $0.2 n_B$ for $\langle : H_1 : \rangle$, $\langle : H_2 : \rangle$ and $\langle : H_{tot} : \rangle_{\rho'}$, respectively.

From these expressions and the first law of entanglement thermodynamics (3.108), entanglement temperature $T_{ent}$ is determined as

$$T_{ent} \simeq \mathcal{N}_T T_{BH}, \tag{3.194}$$

where $T_{BH}$ is the Hawking temperature of the background geometry and $\mathcal{N}_T$ is a constant defined by $\mathcal{N}_T = 2\pi \mathcal{N}_E / \mathcal{N}_S$. Numerical values of $\mathcal{N}_T$ is $\mathcal{N}_T = (n_B - 1/2)\pi/3$, $(n_B + 1/2)\pi/3$, $4 n_B \pi/3$ for $\langle : H_1 : \rangle$, $\langle : H_2 : \rangle$ and $\langle : H_{tot} : \rangle_{\rho'}$, respectively.

These results have several interesting features. First of all we immediately see that the entanglement thermodynamics on the Schwarzschild spacetime shows exactly the same behavior as the black hole thermodynamics. This behavior is just what we expected from the above intuitive argument on gravitational effects: the

[14] The author thanks Professor T. Jacobson for helpful comments on this point.

gravitational redshift effect modifies the area dependence of $E_{ent}$ so as to make the entanglement thermodynamics behave just like the black hole thermodynamics.

Second it should be noted that the temperature $T_{ent}$ becomes independent of the cut-off scale $a$ for all definitions of entanglement energy.

It is also suggestive that the average of entanglement temperatures for $\langle : H_1 : \rangle$ and $\langle : H_2 : \rangle$ with $n_B = 1$ gives almost the same value as $T_{BH}$. This averaging corresponds to averaging out the difference in the red-shift factors for the one-mesh 'inside' and the one-mesh 'outside' of the boundary. Therefore such an averaging may have some meaning.

**Entanglement thermodynamics in Reissner-Nordström spacetime**

From the numerical results in subsection 3.3.3, entanglement entropy and entanglement energy in Reissner-Nordström spacetime are expressed as

$$\begin{aligned} S_{ent} &\simeq k_B \mathcal{N}_S \left(\frac{r_B}{a}\right)^2, \\ E_{ent} &\simeq \hbar c \mathcal{N}_E c(q) \frac{r_B}{a^2}, \end{aligned} \tag{3.195}$$

for all definitions of $E_{ent}$, where $r_B$ is the area radius of the boundary defined by $r_B = r(\rho = n_B a)$. Here $\mathcal{N}_S$ and $\mathcal{N}_E$ are numerical factors of order 1, whose values are the same as those for Schwarzschild spacetime, and the coefficient $c(q)$ is a function of $q = Q/M$ given by Eq. (3.183). The coefficient $c(q)$ approaches to zero in the limit $q \to 1$.

Combining these expressions with the first law of entanglement thermodynamics (3.108), we obtain entanglement temperature $T_{ent}$ as

$$T_{ent} \simeq \mathcal{N}_T T_{BH}, \tag{3.196}$$

where $T_{BH}$ is the Hawking temperature of the background geometry and $\mathcal{N}_T$ is a constant defined by $\mathcal{N}_T = 2\pi \mathcal{N}_E / \mathcal{N}_S$.

Note that both $T_{BH}$ and $T_{ent}$ become zero in the extremal limit ($q \to 1$) of the background spacetime. Therefore as in the case of the Schwarzschild spacetime, the entanglement thermodynamics in the Reissner-Nordström spacetime has the same structure as that of the black-hole thermodynamics.

### 3.3.5 Concluding remark

In this section we have constructed entanglement thermodynamics for a massless scalar field in Minkowski, Schwarzschild and Reissner-Nordström spacetimes. The entanglement thermodynamics in Minkowski spacetime differs significantly from black-hole thermodynamics. On the contrary, the entanglement thermodynamics in Schwarzschild and Reissner-Nordström spacetimes has the same structure as that of black-hole thermodynamics. In particular, it has been shown that entanglement temperature in the Reissner-Nordström spacetime approaches zero in the extremal limit.

Our model analysis strongly suggests a tight connection between the entanglement thermodynamics and the black hole thermodynamics. Of course, our model is too simple to give any definite conclusion based on it. In particular, the ambiguity in the definition of the energy comes from neglecting backreaction of the quantum field on gravity.

Finally we comment on possible extensions of the entanglement thermodynamics. The first is the inclusion of a charged field as matter. In particular, it will be valuable to analyze the entanglement thermodynamics for a charged field in

Reissner-Nordström spacetime. In this case we may be able to define the entanglement charge as an expectation value of the charge of the field for the coarse-grained state. The second is a generalization to non-spherically-symmetric spacetimes. For example it is expected that construction of entanglement thermodynamics in Kerr spacetime requires the introduction of a concept of an entanglement angular-momentum.

## 3.4 A new interpretation of entanglement entropy

In this section a new interpretation of entanglement entropy is proposed: entanglement entropy of a pure state with respect to a division of a Hilbert space into two subspaces 1 and 2 is an amount of information, which can be transmitted through 1 and 2 from a system interacting with 1 to another system interacting with 2. The transmission medium is quantum entanglement between 1 and 2. In order to support the interpretation, suggestive arguments are given: variational principles in entanglement thermodynamics and quantum teleportation. It is shown that a quantum state having maximal entanglement entropy plays an important role in quantum teleportation. Hence, the entanglement entropy is, in some sense, an index of efficiency of quantum teleportation.

In subsection 3.4.1, based on a relation between the entanglement entropy and so-called conditional entropy, we propose an interpretation of the entanglement entropy. In subsection 3.4.2 variational principles in entanglement thermodynamics are used to determine quantum states. In particular, a state having maximal entanglement entropy is determined and is used in subsection 3.4.3 to transmit information about an unknown quantum state. Subsection 3.4.4 is devoted to a summary of this section and to discuss implications for the information loss problem and Hawking radiation.

### 3.4.1 Conditional entropy and entanglement entropy

Entropy plays important roles not only in statistical mechanics but also in information theory. In the latter, entropy of a random experiment, each of whose outcomes has an attached probability, represents uncertainty about the outcome before performing the experiment [87]. Besides the well-known Shannon entropy, there exist various definitions of entropies in information theory. For example, the so-called conditional entropy of an experiment $A$ on another experiment $B$ is defined by $H(A|B) = -\sum_{a,b} p(a,b) \ln p(a|b)$, where $a$ and $b$ represent outcomes of $A$ and $B$, respectively, $p(a,b)$ is a joint probability of $a$ and $b$, and $p(a|b) = p(a,b)/p(b)$ is a conditional probability of $a$ on $b$. Here $p(b)$ is a probability of $b$. The conditional entropy corresponds to an uncertainty about the outcome of $A$ after the experiment $B$ is done. In other words it can be regarded as the amount of information about $A$ which cannot be known from the experiment $B$. The quantum analogue of the conditional entropy was considered in references [88, 89] and is called the von Neumann conditional entropy. Consider a Hilbert space $\mathcal{F}$ of the form (3.109) and let $\rho$ be a density matrix on $\mathcal{F}$. The von Neumann conditional entropy of $\rho$ about the subsystem 1 on the subsystem 2 is defined by

$$S_{1|2} = \mathbf{Tr}\left[\rho\sigma_{1|2}\right], \tag{3.197}$$

where $\sigma_{1|2} = \mathbf{1}_1 \otimes \ln\rho_2 - \ln\rho$. The von Neumann conditional entropy $S_{2|1}$ of $\rho$ about the subsystem 2 on the subsystem 1 is defined in a similar way. It is expected that $S_{1|2}$ (or $S_{2|1}$) represents the amount of the information about the subsystem 1 (or 2) which cannot be known from 2 (or 1, respectively).

The von Neumann conditional entropy can be negative. In fact, it is easy to see that

$$S_{1|2} = S_{2|1} = -S_{ent}, \tag{3.198}$$

if $\rho$ is a pure state. Hence, if $\rho$ is a pure state then the conditional entropy is zero or negative. Our question now is 'what is the meaning of the negative conditional entropy of a pure state?' It might be expected that $|S_{1|2}|$ $(= S_{ent})$ is the amount of the information about 1 (or 2) which can be known from 2 (or 1, respectively). However, this statement is not precise. A precise statement is that it is an amount of information, which can be transmitted through 1 and 2 from a system interacting with 1 to another system interacting with 2. The transmission medium is quantum entanglement between 1 and 2.

The purpose of the remaining part of this section is to give suggestive arguments for this statement.

### 3.4.2 Variational principles in entanglement thermodynamics

In statistical mechanics, the von Neumann entropy is used to determine an equilibrium state: an equilibrium state of an isolated system is determined by maximizing the entropy. Thus, we expect that the entanglement entropy may be used to determine a quantum state.

As an illustration we consider a simple system of two particles, each with spin 1/2: we consider a Hilbert space $\mathcal{F}$ of the form (3.109) and denote an orthonormal basis of $\mathcal{F}_i$ by $\{|\uparrow\rangle_i, |\downarrow\rangle_i\}$ $(i = 1, 2)$. Let $|\phi\rangle$ be an element of $\mathcal{F}$ with unit norm and expand it as

$$|\phi\rangle = a|\uparrow\rangle_1 \otimes |\uparrow\rangle_2 + b|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 + c|\downarrow\rangle_1 \otimes |\uparrow\rangle_2 + d|\downarrow\rangle_1 \otimes |\downarrow\rangle_2, \tag{3.199}$$

where $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$ is understood. The corresponding reduced density matrix is given by

$$\begin{aligned} \rho_2 \quad = \quad & (|a|^2 + |c|^2)|\uparrow\rangle_{22}\langle\uparrow| + (ab^* + cd^*)|\uparrow\rangle_{22}\langle\downarrow| \\ & + (a^*b + c^*d)|\downarrow\rangle_{22}\langle\uparrow| + (|b|^2 + |d|^2)|\downarrow\rangle_{22}\langle\downarrow| \end{aligned} \tag{3.200}$$

and the entanglement entropy can be easily calculated from it. The resulting expression for the entanglement entropy is

$$S_{ent} = -\frac{1+x}{2}\ln\left(\frac{1+x}{2}\right) - \frac{1-x}{2}\ln\left(\frac{1-x}{2}\right), \tag{3.201}$$

where $x = \sqrt{1 - 4|ad - bc|^2}$. By requiring $dS_{ent}/dx = 0$ we obtain the condition $|ad - bc| = 1/2$. Thus a state maximizing the entanglement entropy is

$$|\phi\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2\right) \tag{3.202}$$

up to a unitary transformation in $\mathcal{F}_1$ and the corresponding maximal value of the entanglement entropy is $\ln 2$. This state is well known as the EPR state.

It is notable that the corresponding reduced density matrix $\rho_2$ represents the microcanonical ensemble. This fact is related to the fact that the maximum of entropy gives the microcanonical ensemble in statistical mechanics. Thus, in general, if $\mathcal{F}_1$ and $\mathcal{F}_2$ have the same finite dimension $N$ then a state maximizing the entanglement entropy is written as

$$|\phi\rangle = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left(|n\rangle_1 \otimes |n\rangle_2\right) \tag{3.203}$$

up to a unitary transformation in $\mathcal{F}_1$, where $|n\rangle_1$ and $|n\rangle_2$ $(n = 1, 2, \cdots, N)$ are orthonormal basis of $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. (See Appendix A.6 for a systematic derivation. ) In the next section we use the state (3.203) to transmit information about an unknown quantum state.

In statistical mechanics, free energy $F = E - TS$ can also be used to determine a statistical state: its minimum corresponds to an equilibrium state of a subsystem in contact with a heat bath of temperature $T$, provided that $T$ is fixed. This variational principle in statistical mechanics is based on the following three assumptions.

1. The total system (the subsystem + the heat bath) obeys the principle of maximum of entropy.
2. Total energy (energy of the subsystem + energy of the heat bath) is conserved.
3. The first law of thermodynamics holds for the heat bath.

Is there a corresponding variational principle in the quantum system in the Hilbert space $\mathcal{F}$ of the form (3.109)? The answer is yes. In section 3.3 a concept of entanglement energy has been introduced and a thermodynamical structure, which we call entanglement thermodynamics, has been constructed by using the entanglement entropy and the entanglement energy. Thus we expect that entanglement free energy $F_{ent}$ defined as follows plays an important role in entanglement thermodynamics.

$$F_{ent} = E_{ent} - T_{ent} S_{ent}, \tag{3.204}$$

where $E_{ent}$ is the entanglement energy and $T_{ent}$ is a constant. Among several options, we adopt the second definition (b) of the entanglement energy given in subsection 3.3.2:

$$E_{ent} = \langle : H_2 : \rangle. \tag{3.205}$$

As shown in the following arguments, by minimizing the entanglement free energy, we can obtain a state in $\mathcal{F}$ characterized by the constant $T_{ent}$. Before doing it, here we consider a physical meaning of the principle of minimum of the entanglement free energy. Let us introduce another Hilbert space $\mathcal{F}_{bath}$, which plays the role of the heat bath in the above statistical-mechanical consideration, and decompose it to the direct product $\mathcal{F}_{bath} = \mathcal{F}_{bath1} \otimes \mathcal{F}_{bath2}$. In this situation it is expected that the principle of the minimum entanglement free energy corresponds to the following situation.

1. The total system $\mathcal{F}_{tot} \equiv \mathcal{F} \otimes \mathcal{F}_{bath}$ obeys the principle of maximum of the entanglement entropy with respect to the decomposition $\mathcal{F}_{tot} = \mathcal{F}_{tot1} \otimes \mathcal{F}_{tot2}$, where $\mathcal{F}_{tot1} \equiv \mathcal{F}_1 \otimes \mathcal{F}_{bath1}$ and $\mathcal{F}_{tot2} \equiv \mathcal{F}_2 \otimes \mathcal{F}_{bath2}$.
2. Total entanglement energy (entanglement energy for $\mathcal{F}$ + entanglement energy for $\mathcal{F}_{bath}$) is conserved.
3. The first law of entanglement thermodynamics (3.108) holds for $\mathcal{F}_{bath}$. In this situation we call the constant $T_{ent}$ the entanglement temperature.

It must be mentioned here that the variational principle of minimum of the entanglement free energy is not as fundamental as the principle of maximum of the entanglement entropy but is an approximation to the latter principle for a large system. However, like the principle of minimum free energy in statistical mechanics, the former principle should be a very useful tool to determine a quantum state.

We now calculate $F_{ent}$ for the system of two spin-1/2 particles and minimize it. For simplicity we adopt the following Hamiltonian for the subsystem 2:

$$\begin{aligned} {}_2\langle\uparrow | : H_2 : |\uparrow\rangle_2 &= \epsilon, \\ {}_2\langle\uparrow | : H_2 : |\downarrow\rangle_2 &= 0, \\ {}_2\langle\downarrow | : H_2 : |\downarrow\rangle_2 &= 0, \end{aligned} \tag{3.206}$$

where $\epsilon$ is a positive constant. The entanglement free energy $F_{ent}$ for the state (3.199) is given by

$$F_{ent} = \epsilon(|a|^2 + |c|^2) + T_{ent}\left[\frac{1+x}{2}\ln\left(\frac{1+x}{2}\right) + \frac{1-x}{2}\ln\left(\frac{1-x}{2}\right)\right]. \quad (3.207)$$

By minimizing it we obtain the following expression for the state $|\phi\rangle$ up to a unitary transformation in $\mathcal{F}_1$.

$$|\phi\rangle = \frac{1}{\sqrt{Z}}\left[e^{-\epsilon/2T_{ent}}|\uparrow\rangle_1 \otimes |\uparrow\rangle_2 + |\downarrow\rangle_1 \otimes |\downarrow\rangle_2\right], \quad (3.208)$$

where $Z = e^{-\epsilon/T_{ent}} + 1$.

The corresponding reduced density matrix $\rho_2$ on $\mathcal{F}_2$ represents a canonical ensemble with temperature $T_{ent}$. This fact is related to the fact that the principle of minimum of free energy results in a canonical ensemble in statistical mechanics. Thus, in general, if $\mathcal{F}_1$ and $\mathcal{F}_2$ have the same finite dimension $N$ then a state minimizing the entanglement free energy is written as

$$|\phi\rangle = \frac{1}{\sqrt{Z}}\sum_{n=1}^{N}\left[e^{-E_n/2T_{ent}}|n\rangle_1 \otimes |n\rangle_2\right] \quad (3.209)$$

up to a unitary transformation in $\mathcal{F}_1$, where $Z = \sum_{n=1}^{N} e^{-E_n/T_{ent}}$, and $E_n$ and $|n\rangle_2$ $(n = 1, 2, \cdots, N)$ are eigenvalues and orthonormalized eigenstates of the normal-ordered Hamiltonian of the subsystem 2. (See Appendix A.6 for a systematic derivation.)

The state (3.209) can be obtained also from another version of the principle of maximum of the entanglement entropy: if we maximize $S_{ent}$ with $E_{ent}$ fixed then the state (3.209) is obtained. In this case, the constant $T_{ent}$ is determined so that the entanglement energy coincides with the fixed value.

Note that in Eq. (3.209) the infinite-dimensional limit $N \to \infty$ can be taken, provided that $T_{ent}$ is bounded. In this limit, the state (3.209) has the same form as those appearing in the thermo field dynamics of black holes [85] and the quantum field theory on a collapsing star background [93]. In fact, if we can set the value of the entanglement temperature of $\mathcal{F}_{bath}$ to be the black hole temperature then the state (3.209) in the limit completely coincides with those in Refs. [85, 93]. In section 3.3 it has been shown numerically that the entanglement temperature for a real massless scalar field in Schwarzschild and Reissner-Nordström spacetimes is finite and equal to the black hole temperature of the background geometry up to a numerical constant of order 1. The finiteness of the entanglement temperature in the black hole spacetimes is a result of cancellation of divergences in entanglement entropy and entanglement energy [46]. Thus, the finiteness is preserved even in the limit of zero cutoff length ($a \to 0$).

### 3.4.3 Quantum teleportation

In Ref. [90] Bennet et al. proposed a method of teleportation of an unknown quantum state from one place to another. It is called quantum teleportation. In their method the information about the quantum state is separated into a 'quantum channel' and a 'classical channel', and each channel is sent separately from a sender "Alice" to a receiver "Bob". What is important is that the quantum channel is sent in a superluminal way by using a quantum correlation or entanglement, while the classical channel is transmitted at most in the speed of light. Here we mention that causality is not violated in an informational sense since Bob cannot obtain any useful information about the unknown state before the arrival of the classical channel.

Hence Alice has to deliver the classical channel to Bob without fail. On the contrary she does not need to worry about whether the information in the quantum channel arrives at Bob's hand since the arrival is guaranteed by the quantum mechanics. It is notable that recently quantum teleportation was confirmed by experiments [91, 92].

In this section we generalize the arguments in Ref. [90] to more abundant situations and try to reformulate it in terms of the entanglement entropy.

Let us consider a Hilbert space $\mathcal{F}$ of the form (3.109) with $\mathcal{F}_i$ constructed from Hilbert spaces $\mathcal{F}_{i\pm}$ as

$$\mathcal{F}_i = \mathcal{F}_{i+} \otimes \mathcal{F}_{i-}. \tag{3.210}$$

For example, consider matter fields in a black hole spacetime formed by gravitational collapse. In this situation, let $\mathcal{H}_1$ be a space of all wave packets on the future event horizon and $\mathcal{H}_2$ be a space of all wave packets on the future null infinity, and decompose each $\mathcal{H}_i$ into a high frequency part $\mathcal{H}_{i+}$ and a low frequency part $\mathcal{H}_{i-}$. Typically, we suppose the decomposition at an energy scale of Planck order. If we define $\mathcal{F}_{i\pm}$ as Fock spaces constructed from $\mathcal{H}_{i\pm}$, respectively, then the space $\mathcal{F}$ of all quantum states of the matter fields is given by (3.109) with (3.210). Although the following arguments do not depend on the construction of the Hilbert space $\mathcal{F}$, this example should be helpful for us to understand the physical meaning of the results obtained.

For simplicity we consider the case that all $\mathcal{F}_{i\pm}$ have the same finite dimension $N$ although in the above example of field theory the dimensions of the Hilbert spaces are infinite [15]. The main purpose of this section is to see general properties of the entanglement entropy by using a finite system. Anyway, the above example of field theory may be helpful in understanding the following arguments. In the finite dimensional case we assume the following three physical principles.

(a) A quantum state $|\phi\rangle$ in $\mathcal{F}$ is a direct product state given by $|\phi_+\rangle_+ \otimes |\phi_-\rangle_-$, where $|\phi_\pm\rangle_\pm$ are elements of $\mathcal{F}_\pm = \mathcal{F}_{1\pm} \otimes \mathcal{F}_{2\pm}$, respectively.

(b) $|\phi_+\rangle_+$ is determined by the principle of maximum of the entanglement entropy with respect to the decomposition $\mathcal{F}_+ = \mathcal{F}_{1+} \otimes \mathcal{F}_{2+}$.

(c) A complete measurement of the von Neumann type on the joint system $\mathcal{F}_1$ is performed by a sender (Alice) in the orthonormal basis $\{|\psi_{nm}\rangle_1\}$, each of which maximizes the entanglement entropy with respect to the decomposition $\mathcal{F}_1 = \mathcal{F}_{1+} \otimes \mathcal{F}_{1-}$.

In other words the assumption (c) is stated as follows: the state $|\phi\rangle$ is projected by one of the basis $|\psi_{nm}\rangle_1$.

In the following arguments, under these assumptions, we show a possibility of quantum teleportation of the state $|\phi_-\rangle_-$ in $\mathcal{F}_-$ to $\mathcal{F}_2$: we make a clone of $|\phi_-\rangle_-$ by using the quantum entanglement which the state $|\phi_+\rangle_+$ has. Therefore a receiver (Bob), who cannot contact with $\mathcal{F}_1$, may be able to get all information about the state $|\phi_-\rangle_-$ in $\mathcal{F}_-$, provided that he can manage to get the classical channel.

Now let us show that explicitly. By the assumption (b) and the arguments in subsection 3.4.2 (see Eq. (3.203)), the state $|\phi_+\rangle_+$ can be written as

$$|\phi_+\rangle_+ = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} |n\rangle_{1+} \otimes |n\rangle_{2+}, \tag{3.211}$$

[15] In applying the results for finite dimensions to field theory, we have to introduce a regularization scheme to make the system finite. For example, we can discretize the system by introducing a cutoff length. After that, we can consider a finite dimensional subspace of the total Hilbert space of the discretized theory, for example, by restricting total energy to be less than the mass of the background geometry. After performing all calculations, we have to confirm that the infinite-dimensional limit can be taken. See, for example, the final paragraph of the previous subsection.

where $\{|n\rangle_{i+}\}$ $(n = 1, 2, \cdots, N)$ are an orthonormal basis of $\mathcal{F}_{i+}$. Next, expand $|\phi_-\rangle_-$ as

$$|\phi_-\rangle_- = \sum_{nm} C_{nm}|n\rangle_{1-} \otimes |m\rangle_{2-}, \tag{3.212}$$

where $\{|n\rangle_{i-}\}$ $(n = 1, 2, \cdots, N)$ are an orthonormal basis of $\mathcal{F}_{i-}$, and $\sum_{nm} |C_{nm}|^2 = 1$ is understood. To impose the assumption (c), we adopt the following basis $\{|\psi_{nm}\rangle_1\}$ $(n, m = 1, 2, \cdots, N)$, each of which maximizes the entanglement entropy.

$$|\psi_{nm}\rangle_1 = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} e^{2\pi ijn/N} |(j+m) mod N\rangle_{1+} \otimes |j\rangle_{1-}. \tag{3.213}$$

In Appendix A.7 it is proved that (3.213) is unique up to a unitary transformation in $\mathcal{F}_{1+}$. Hence, $|\phi\rangle = |\phi_+\rangle_+ \otimes |\phi_-\rangle_-$ is written as

$$|\phi\rangle = \frac{1}{N} \sum_{nm} |\psi_{nm}\rangle_1 \otimes U_{nm}^{(2+)} |\tilde{\phi}_2\rangle_2, \tag{3.214}$$

where $|\tilde{\phi}_2\rangle_2$ is a state in $\mathcal{F}_2$ given by

$$|\tilde{\phi}_2\rangle_2 = \sum_{n'm'} C_{n'm'} |n'\rangle_{2+} \otimes |m'\rangle_{2-}, \tag{3.215}$$

and $U_{nm}^{(2+)}$ $(n, m = 1, 2, \cdots, N)$ are unitary transformations in $\mathcal{F}_{2+}$ defined by

$$U_{nm}^{(2+)} = \sum_{k=1}^{N} e^{-2\pi ikn/N} |(k+m) mod N\rangle_{2+2+}\langle k|. \tag{3.216}$$

(See Appendix A.7 for an explicit derivation of (3.214).)

Thus, after the measurements in the basis $\{|\psi_{nm}\rangle_1\}$ by the sender (Alice), the original state $|\phi\rangle$ jumps to one of the states $|\tilde{\phi}_{nm}\rangle$ defined by

$$|\tilde{\phi}_{nm}\rangle = |\psi_{nm}\rangle_1 \otimes U_{nm}^{(2+)} |\tilde{\phi}_2\rangle_2. \tag{3.217}$$

This state can be seen by the receiver (Bob), who cannot contact with $\mathcal{F}_1$, as the state $U_{nm}^{(2+)}|\tilde{\phi}_2\rangle_2$ in $\mathcal{F}_2$. Here note that the unitary transformation $U_{nm}^{(2+)}$ in $\mathcal{F}_{2+}$ is completely determined by a pair of integers $n$ and $m$ (outcome of the experiment by Alice). Thus, if the two integers are sent to the receiver (Bob) in the classical channel, then by operating the inverse transformation of the corresponding unitary transformation in $\mathcal{F}_{2+}$ the receiver (Bob) can obtain the 'clone' state $|\tilde{\phi}_2\rangle_2$ $(\in \mathcal{F}_2)$ of $|\phi_-\rangle$ $(\in \mathcal{F}_-)$. It is evident that $|\tilde{\phi}_2\rangle_2$ has all information about the original state $|\phi_-\rangle$.

It is remarkable that information to be sent to the receiver (Bob) in the classical channel is only two integers $n$ and $m$, while information included in the unknown state $|\phi_-\rangle_-$ is a set of complex constants $\{C_{nm}\}$ $(n, m = 1, 2, \cdots, N)$ with a constraint $\sum_{nm} |C_{nm}|^2 = 1$. Thus a large amount of information is sent in the quantum channel. Here we mention that tracing out $\mathcal{F}_{2+}$ from the state $U_{nm}^{(2+)}|\tilde{\phi}_2\rangle_2$ or $|\tilde{\phi}_2\rangle_2$ results in the following density matrix $\rho_{2-}$ on $\mathcal{F}_{2-}$:

$$\rho_{2-} = \sum_{nm} \left( \sum_j C_{jn} C_{jm}^* \right) |n\rangle_{2-2-}\langle m|, \tag{3.218}$$

which is equivalent to the density matrix obtained by tracing out $\mathcal{F}_{1-}$ from the original unknown state $|\phi_-\rangle_-$. Hence, if the receiver (Bob) cannot contact with $\mathcal{F}_{2+}$, he does not obtain any information from the sender (Alice).

Finally it must be mentioned that the success of quantum teleportation is due to quantum entanglement in the state $|\phi_+\rangle_+$ which has maximal entanglement entropy. If we took $|\phi_+\rangle_+$ with less entanglement entropy then the teleportation would be less successful. Therefore, the entanglement entropy can be regarded as an index of efficiency of quantum teleportation. This consideration supports the interpretation of the entanglement entropy proposed in subsection 3.4.1.

### 3.4.4 Concluding remark and physical implications

In this section a new interpretation of entanglement entropy has been proposed based on its relation to the so-called conditional entropy and a well-known meaning of the latter. It is conjectured that entanglement entropy of a pure state with respect to a division of a Hilbert space into two subspaces 1 and 2 is an amount of information, which can be transmitted through 1 and 2 from a system interacting with 1 to another system interacting with 2. The medium of the transmission is quantum entanglement between 1 and 2.

To support the interpretation we have given the following two suggestive arguments: variational principles in entanglement thermodynamics and quantum teleportation. The most important variational principle we considered is the principle of maximum of entanglement entropy. This principle determines a state uniquely up to a unitary transformation in one of the two Hilbert subspaces (not in the whole Hilbert space). From the proposed conjecture it is expected that information can be transmitted most effectively through the two subspaces by using the maximal entanglement of the state. In fact, reformulating the quantum teleportation in terms of the entanglement entropy, we have shown that the state having maximal entanglement entropy plays an important role in quantum teleportation. This consideration gives strong support to our interpretation.

As a by-product we have shown that the variational principle of minimum of entanglement free energy is useful to determine a quantum state. The resulting quantum state has exactly the same form as those appearing in the thermo field dynamics of black holes [85] and the quantum field theory on a collapsing black hole background [93], provided that the entanglement temperature $T_{ent}$ is set to be the black hole temperature. It is remarkable that, as shown in section 3.3, $T_{ent}$ for a real massless scalar field in Schwarzschild and Reissner-Nordström spacetimes is equal to the black hole temperature of the background geometry up to a numerical constant of order 1. Thus we can say that the variational principle of minimum of entanglement free energy gives a new derivation of the Hawking radiation. Finally, we mention that with this variational principle the entanglement thermodynamics is equivalent to 'tHooft's brick wall model [35].

It will be valuable to analyze how to generalize arguments in this section to the situation that divergences in entropy and energy are absorbed by renormalization [76, 94]. If the generalization is achieved, the physical meaning of the entanglement entropy in black hole physics will become clearer. It is noteworthy that in the brick wall model, as shown in section 3.2, the divergence in thermal energy is exactly canceled by divergence in negative energy [44].

Now the final comment is in order. It is worthwhile to clarify in what physical situations the variational principles can be applicable. (In thermodynamics the second law supports the principle of maximum entropy.) In other words, in what situations does the entanglement entropy increase? In what situations does the entanglement free energy decrease? To answer these questions, theorem 7 in subsection 2.3.2 or its generalization may be useful.

# Chapter 4

# Discussions

In this thesis we have analyzed properties and the origin of the black hole entropy in detail from various points of view.

First, in chapter 2 laws of black hole thermodynamics have been reviewed. In particular, the first and generalized second laws have been investigated in detail. It is in these laws that the black hole entropy plays key roles.

In section 2.1 we have re-analyzed Wald and Iyer-Wald derivation of the first law of black hole mechanics in a general covariant theory of gravity, following Ref. [40]. In particular, two issues listed in the beginning of section 2.1 have been discussed in detail: (a) gauge conditions and (b) near-stationary black hole entropy. It has been shown that the minimal set of gauge conditions necessary for the derivation of the first law for stationary black holes is that $t^a$ and $\varphi^a$ are fixed at spatial infinity, where $t^a$ is the stationary Killing field with unit norm at infinity, and $\varphi^a$ denotes axial Killing fields. It has also been shown that for non-stationary perturbations about a stationary perturbation the first law does hold to first order in perturbation.

However, this first law cannot be applied to a purely dynamical situation. In this sence we have called it the first law of black hole statics. The purpose of section 2.2 has been to consider dynamical definition of black hole entropy and to derive a dynamical version of the first law of black hole, which we call the quasi-local first law of black hole dynamics. For simplicity, we have considered the general relativity only. Extension to a general covariant theory of gravity will be valuable.

In subsection 2.2.1 we have considered two non-statistical definitions of entropy for dynamic (non-stationary) black holes in spherical symmetry. The first is analogous to the original Clausius definition of thermodynamic entropy: there is a first law containing an energy-supply term which equals surface gravity times a total differential. The second is Wald's Noether-charge method, adapted to dynamic black holes by using the Kodama flow. It has been shown that both definitions give the same answer for Einstein gravity: one-quarter the area of the trapping horizon [41].

In subsection 2.2.1 the quasi-local first law of black hole dynamics has been derived without assuming any symmetry and any asymptotic condition [42]. In the derivation we have given a new definition of dynamical surface gravity. In spherical symmetry it reduces to that defined in Ref. [55].

In section 2.3 we have proved the generalized second law for a quasi-stationary black hole which is formed by gravitational collapse [10]. After that, in subsection 2.3.3, we have discussed a generalization of our proof to a dynamical background. It has been suggested that the generalization may be possible by using the quasi-local first law derived in subsection 2.2.1.

Next, in chapter 3 three candidates for the origin of the black hole entropy have been analyzed: the D-brane statistical-mechanics, the brick wall model, and the entanglement thermodynamics.

In section 3.1 the D-brane statistical-mechanics has been reviewed by using a configuration of D-strings and D-fivebranes wrapped on $T^5 = T^4 \times S^1$, which was introduced in Ref. [43]. We have consider a set of multiply-wound D-strings, which is composed of $N_{q_1}^{(1)}$ D-strings of length $2\pi R q_1$ ($q_1 = 1, 2, \cdots$) along the $S^1$ and a set of multiply-wound D-fivebranes, which is composed of $N_{q_5}^{(5)}$ D-fivebranes of length $2\pi R q_5$ ($q_5 = 1, 2, \cdots$) along the $S^1$. Here $\{N_{q_1}^{(1)}\}$ and $\{N_{q_5}^{(5)}\}$ are sets of arbitrary non-negative integers, and $R$ is radius of the $S^1$. It has been shown that the number of microscopic states of open strings on the D-branes is bounded from above by exponential of the Bekenstein-Hawking entropy of the corresponding black hole, and that the temperature of a decay of D-brane excitations to closed strings is bounded from below by the Hawking temperature of the corresponding black hole. This result has been summarized as Eqs.(3.31) and (3.37). The necessary and sufficient condition for these bounds to be saturated has been shown explicitly and some speculations has been given.

In section 3.2 we have re-examined the brick wall model to solve problems concerning this model. In particular, it has been shown that the wall contribution to the total gravitational mass is zero if and only if temperature of thermal gas measured at infinity is set to be the Hawking temperature, and that the backreaction can be neglected [44].

In section 3.3 we have constructed entanglement thermodynamics for a massless scalar field in Minkowski [45], Schwarzschild [46] and Reissner-Nordström spacetimes. The entanglement thermodynamics in Minkowski spacetime differs significantly from black-hole thermodynamics. On the contrary, the entanglement thermodynamics in Schwarzschild and Reissner-Nordström spacetimes has the same structure as that of black-hole thermodynamics. In particular, it has been shown that entanglement temperature in the Reissner-Nordström spacetime approaches zero in the extremal limit.

In section 3.4 a new interpretation of entanglement entropy has been proposed based on its relation to the so-called conditional entropy and a well-known meaning of the latter [47]. It has been conjectured that entanglement entropy of a pure state with respect to a division of a Hilbert space into two subspaces 1 and 2 is an amount of information, which can be transmitted through 1 and 2 from a system interacting with 1 to another system interacting with 2. The medium of the transmission is quantum entanglement between 1 and 2. To support the interpretation we have given the following two suggestive arguments: variational principles in entanglement thermodynamics and quantum teleportation. It has been shown that the state having maximal entanglement entropy plays an important role in quantum teleportation. This consideration gives strong support to our interpretation.

Now let us discuss about semiclassical consistencies of the brick wall model and the entanglement thermodynamics.

It has been shown that the brick wall model is a consistent semiclassical description of black hole entropy: thermal excitations raised to the Hawking temperature above the Boulware state explains the black hole entropy; the positive divergence in thermal energy is canceled by the negative divergence in the vacuum energy of the Boulware state. Namely, the following simultaneous equations have a solution:

$$\begin{aligned} S_{wall} &= S_{BH}, \\ T_\infty &= T_{BH}, \\ (\Delta M)_{therm,wall} + (\Delta M)_{B,wall} &= 0. \end{aligned} \tag{4.1}$$

It has been found that the last equation is equivalent to the second one. Thus, the

solution is

$$\begin{aligned} \alpha &= \sqrt{\frac{\mathcal{N}}{90\pi}} l_{pl}, \\ T_\infty &= T_{BH}. \end{aligned} \tag{4.2}$$

On the other hand, entanglement thermodynamics has the following properties common to all definitions of entanglement energy.

$$\begin{aligned} S_{ent} &\simeq N\mathcal{N}_S \left(\frac{r_0}{a}\right)^2, \\ E_{ent} &\simeq N\mathcal{N}_E \frac{r_0}{a^2}, \\ T_{ent} &\simeq \mathcal{N}_T T_{BH}, \end{aligned} \tag{4.3}$$

where $\mathcal{N}_S$ and $\mathcal{N}_E$ are numerical factors of order unity, and $\mathcal{N}_T = 2\pi\mathcal{N}_E/\mathcal{N}_S$. Here, we multiplied entanglement entropy $S_{ent}$ and energy $E_{ent}$ by the number $N$ of fields [1]. These properties are expected to hold also for states different from the Boulware state, provided that the definition of entanglement quantities are properly modified. For other states, of course, the numerical factors $\mathcal{N}_S$ and $\mathcal{N}_E$ will be different from those for the Boulware state.

Our question now is whether the entanglement thermodynamics is a consistent semiclassical description of black hole thermodynamics or not. Here let us assume that the sum of the vacuum energy and the entanglement energy contributes to gravitational energy. Hence, the question is restated as whether the following simultaneous equations hold or not.

$$\begin{aligned} S_{ent} &= S_{BH}, \\ T_{ent} &= T_{BH}, \\ E_{ent} + (\Delta M)_{B,wall} &= 0. \end{aligned} \tag{4.4}$$

Since the cutoff length $\alpha$ in section 3.2 seems to be related to the cutoff length $a$ in section 3.3 as

$$\alpha \simeq n_B a, \tag{4.5}$$

a solution of Eqs. (4.4) is given by

$$\begin{aligned} a &\simeq \sqrt{\frac{N\mathcal{N}_S}{\pi}} l_{pl}, \\ \mathcal{N}_E &\simeq \frac{\mathcal{N}_S}{2\pi}, \\ n_B &\simeq \sqrt{\frac{\pi^4}{90} \cdot \frac{1}{240\mathcal{N}_S}}. \end{aligned} \tag{4.6}$$

Unfortunately, since $\mathcal{N}_S = 0.3$, the r.h.s of the last equation is less then unity. Hence the last equation does not hold for positive integer value of $n_B$. Even if $n_B$ would be allowed to be less than unity, the value of $n_B$ given by the last equation would not be consistent with the second equation for all our definitions of entanglement energy [2]. This discrepancy suggests that our model of entanglement thermodynamics suffers from a strong backreaction near horizon. This might mean that our choice of the pure state (the Boulware state) would be wrong. However, we strongly expect that there is a state for which the consistency condition (4.6) holds. It will be worthwhile to analyze whether Eqs. (4.6) are satisfied for the Hartle-Hawking state or not.

[1] Note that $N = 1$ corresponds to one bosonic field. Hence, $\mathcal{N} = N\pi^4/90$.
[2] Note that $\mathcal{N}_E$ depends on $n_B$.

Next, we shall combine a variational principle in entanglement thermodynamics with the third equation in the semiclassical consistency conditions. As explained in subsection 3.4.2, the principle of maximum of entanglement entropy introduced in Ref. [47] determines a quantum state to be (3.203) up to a unitary transformation in one of two Hilbert subspaces. Moreover, the principle of minimum of entanglement free energy determines a quantum state to be (3.209) up to a unitary transformation in one of two subspaces, too. It has also been mentioned in the second-to-last paragraph of subsection 3.4.2 that, if we maximize entanglement entropy with entanglement energy fixed, the state (3.209) is obtained. In this case, entanglement temperature should be determined so that the entanglement energy coincides with the fixed value. On the other hand, the third equation in the semiclassical consistency conditions (4.4) does fix entanglement energy, provided that we do not change $(\Delta M)_{B,wall}$. We shall call this condition the small backreaction condition. Hence, the principle of maximum of entanglement entropy combined with the small backreaction condition determines a state of a quantum field near horizon to be (3.209) up to a unitary transformation in one of two Hilbert subspaces, provided that a suitable regularization scheme is introduced. This state has exactly the same form as those appearing in the thermo field dynamics of black holes [85] and the quantum field theory on a collapsing black hole background [93], provided that the entanglement temperature $T_{ent}$ can be set to be the Hawking temperature. In this case, since the entanglement energy is the same as thermal energy in the brick wall model because of the small backreaction condition, and since the entanglement entropy is expected to have the almost same value as the Bekenstein-Hawking entropy, the entanglement temperature also seems to coincide with the Hawking temperature up to a numerical factor of order unity. Therefore, it seems that the principle of maximum of entanglement entropy combined with the small backreaction condition may provide a new derivation of the Hawking radiation. Further investigations on this point will be valuable.

Finally, let us discuss the information loss problem.

We have proposed a new interpretation of entanglement entropy: entanglement entropy of a pure state with respect to a division of a Hilbert space into two subspaces 1 and 2 is an amount of information, which can be transmitted through 1 and 2 from a system interacting with 1 to another system interacting with 2. On the other hand, it has been confirmed in section 3.3 and in many references [37, 38, 39, 46] that the entanglement entropy has the same value as the black hole entropy up to a numerical constant of order unity, provided that a cutoff length of Planck order is introduced in the theory. Hence we have a large amount of entanglement entropy to transmit information from inside to outside of a black hole by using quantum entanglement.

Hence, in our interpretation, it seems that the entanglement entropy is a quantity which cancels the black hole entropy to restore information loss, provided that the black hole entropy represents the amount of the information loss. For example, suppose that a black hole is formed from an initial state with zero entropy ($S = 0$). In this case, non-zero black hole entropy is generated ($S_{BH} > 0$) from the zero entropy state. At the same time, entanglement entropy and negative conditional entropy are also generated and their absolute values are as large as the black hole entropy ($S_{ent} = |S_{cond}| \simeq S_{BH}$). After that, the black hole evolves by emitting Hawking radiation, changing the value of $S_{BH}$ and $S_{ent}$ ($= |S_{cond}|$) with $S_{BH} \simeq S_{ent}$ kept. Finally, when the black hole evaporates, the entanglement entropy cancels the black hole entropy to settle the final entropy to be zero ($S = 0$). To summarize, the black hole entropy is an amount of temporarily missing information and the entanglement entropy is a quantity which cancels the black hole entropy. Both entropies appear and disappear together from the sea of zero entropy state.

We conclude this thesis by giving some speculations.

(i) The black hole entropy should be related to a number of microscopic states in the corresponding black hole. The microscopic description should be possible by using a quantum theory of gravity (eg. superstring theory, loop quantum gravity, etc.).

(ii) The existence of the horizon prevents the microscopic states in the black hole from being seen on the outside. Hence, information about the microscopic states is lost at least temporarily. (See Ref. [95] for Hawking's objection to string-theorist's point of view.) At the same time, the horizon generates entanglement entropy of matter fields which has the same value as the black hole entropy.

(iii) The entanglement entropy is an amount of information, which can be transmitted through the matter fields from a system interacting with the matter fields in the black hole to another system interacting with the matter fields outside the black hole, provided that the classical channel in the sense of the quantum teleportation can be transmitted properly. The former system carries the temporarily missing information corresponding to the black hole entropy. Hence the quantum channel of all temporarily missing information about the microscopic states can be transmitted to the outside of the black hole before the black hole evaporates completely. It is the classical channel that is necessary for restoring the original information from the quantum channel.

(iv) We propose the conjecture that one of the following two should be realized.

  (a) The black hole evaporates completely, and the classical channel is transmitted at the final stage of the black hole evaporation.

  (b) There remains a remnant at the end of the Hawking radiation, and all or a part of the classical channel is carried by the remnant forever.

  If (a) is correct, then it is in principle possible to restore all information about the microscopic states of the black hole after the evaporation. If (b) is correct, then all or a part of the information about the microscopic states cannot be restored although all the information remain to exist: the information cannot be decoded since the classical and quantum channels are located separately.

# Acknowledgments

I would like to thank my advisor, Professor H. Kodama, for his letting me know how to proceed research works and encouraging me continuously. I would like to express my appreciation to Professor W. Israel for helpful discussions and continuing encouragement during and after his stay in Kyoto as a visiting professor. I am very grateful to M. Seriu for many helpful discussions and constructive suggestions. I would like to acknowledge stimulating discussions with M. Siino, T. Chiba, K. Nakao and S. A. Hayward, and with other colleagues in Yukawa Institute for Theoretical Physics and Department of Physics in Kyoto University. Finally, I am grateful to my wife, Kyoko.

# Appendix

## A.1 The conditional probability

In this appendix we reduce (2.62) to (2.64). First the $S$-matrix obtained by [67] is

$$\begin{aligned} S|0\rangle &= N\sum_{n=0}^{\infty}\frac{\sqrt{(2n)!}}{2^n n!}\left(\overset{n}{\otimes}\epsilon\right)_{sym}, \\ Sa^\dagger({}_i\gamma)S^{-1} &= R_i a^\dagger({}_i\rho)+T_i a^\dagger({}_i\sigma), \end{aligned} \tag{A.7}$$

where $\epsilon$ and $N$ are a bivector and a normalization constant defined by

$$\epsilon = 2\sum_i x_i({}_i\lambda\otimes{}_i\tau)_{sym}, \quad N=\prod_i\sqrt{1-x_i},$$

where

$$x_i = \exp\left(-\pi(\omega_i-\Omega_{BH}m_i)/\kappa\right).$$

In the expression, $\omega_i$ and $m_i$ are a frequency and an azimuthal angular momentum quantum number of a mode specified by integer $i$, $\Omega_{BH}$ and $\kappa$ are an angular velocity and a surface gravity of the black hole. ${}_i\gamma, {}_i\rho, {}_i\sigma, {}_i\lambda$ and ${}_i\tau$ are unit vectors in $\mathcal{H}_{\mathcal{I}^+}\oplus\mathcal{H}_{H^+}$ defined in [67], and the former four are related as follows:

$$\begin{aligned} {}_i\gamma^a &= T_i\ {}_i\sigma^a + R_i\ {}_i\rho^a, \\ {}_i\lambda^a &= t_i\ {}_i\rho^a + r_i\ {}_i\sigma^a, \end{aligned} \tag{A.8}$$

where $t_i$, $T_i$ are transmission coefficients for the mode specified by the integer $i$ on the Schwarzschild metric [67] and $r_i$, $R_i$ are reflection coefficients. They satisfy [3]

$$\begin{aligned} |t_i|^2+|r_i|^2 &= |T_i|^2+|R_i|^2=1, \\ t_i=T_i, &\qquad r_i=-R_i^*T_i/T_i^*. \end{aligned} \tag{A.9}$$

By using the S-matrix, we obtain

$$\begin{aligned} S|\{n_{i\gamma}\}\rangle &= N\left[\prod_i\frac{1}{\sqrt{n_{i\gamma}!}}\left[R_i a^\dagger({}_i\rho)+T_i a^\dagger({}_i\sigma)\right]^{n_{i\gamma}}\right]\sum_{n=0}^{\infty}\frac{\sqrt{(2n)!}}{2^n n!}\left(\overset{n}{\otimes}\epsilon\right)_{sym} \\ &= N\sum_{n=0}^{\infty}{\sum_{}}'\left[\prod_i\frac{1}{\sqrt{n_{i\gamma}!}}\begin{pmatrix} n_{i\gamma} \\ m_i\end{pmatrix}R_i^{m_i}T_i^{n_{i\gamma}-m_i}\right] \\ &\quad\times\sqrt{(2n)!}\left[\prod_i\frac{x_i^{n_i}}{n_i!}\begin{pmatrix} n_i \\ l_i\end{pmatrix}t_i^{l_i}r_i^{n_i-l_i}\right] \end{aligned}$$

[3] The last two equations are consequences of the time reflection symmetry of the Schwarzschild metric.

$$
\begin{aligned}
&\times\sqrt{\frac{(2n+\sum_i n_{i\gamma})!}{(2n)!}}\left(\prod_i {}^{n_i}_{\otimes}{}_i\tau^{\otimes\, l_i+m_i}{}_i\rho^{\otimes\, n_i-l_i+n_{i\gamma}-m_i}{}_i\sigma\right)_{sim}\\
=\;& N\sum_{n_i=0}^{\infty}\sum_{m_i=0}^{n_{i\gamma}}\sum_{l_i=0}^{n_i}\sqrt{\left(\sum_i(2n_i+n_{i\gamma})\right)!}\\
&\times\prod_i\left[\frac{1}{\sqrt{n_{i\gamma}!}}\cdot\frac{x_i^{n_i}}{n_i!}\begin{pmatrix} n_{i\gamma}\\ m_i\end{pmatrix}\begin{pmatrix} n_i\\ l_i\end{pmatrix}R_i^{m_i}T_i^{n_{i\gamma}-m_i}t_i^{l_i}r_i^{n_i-l_i}\right]\\
&\times\left(\prod_i {}^{n_i}_{\otimes}{}_i\tau^{\otimes\, l_i+m_i}{}_i\rho^{\otimes\, n_i-l_i+n_{i\gamma}-m_i}{}_i\sigma\right)_{sym},
\end{aligned}
\tag{A.10}
$$

where $\sum'$ denotes a summation with respect to $n_i$, $m_i$ and $l_i$ over the following range: $\sum_i n_i = n$, $n_i \geq 0$, $0 \leq m_i \leq n_{i\gamma}$, $0 \leq l_i \leq n_i$. In (A.10), those orthonormal basis vectors in $|\{n_{i\rho}\}\rangle \otimes \mathcal{F}(\mathcal{H}_{H^+})$ that have a non-vanishing inner product with $S|\{n_{i\gamma}\}\rangle$ appear in the form [4]

$$
\sqrt{\frac{(\sum_i(2n_i+n_{i\gamma}))!}{\prod_i[n_i!n_{i\rho}!(n_i+n_{i\gamma}-n_{i\rho})!]}}\left(\prod_i {}^{n_i}_{\otimes}{}_i\tau^{\otimes\, n_{i\rho}}{}_i\rho^{\otimes\, n_i+n_{i\gamma}-n_{i\rho}}{}_i\sigma\right)_{sym}. \tag{A.11}
$$

Thus, when we calculate $(\langle\{n_{i\rho}\}| \otimes \langle H|)\, S|\{n_{i\gamma}\}\rangle$, the summation in (A.10) is reduced to a summation with respect to $n_i$ and $m_i$ over the range $n_i \geq \max(0, n_{i\rho} - n_{i\gamma})$, $\max(0, n_{i\rho} - n_i) \leq m_i \leq \min(n_{i\rho}, n_{i\gamma})$ with $l_i = n_{i\rho} - m_i$ [5]. Here $|H\rangle$ is an element of $\mathcal{F}(\mathcal{H}_{H^+})$. Paying attention to this fact, we can obtain the following expression of the conditional probability.

$$
\begin{aligned}
&P(\{n_{i\rho}\}|\{n_{i\gamma}\})\\
&= |N|^2\sum_{\{n_i\geq\max(0,n_{i\rho}-n_{i\gamma})\}}\left(\sum_i(2n_i+n_{i\gamma})\right)!\\
&\times\prod_i\left[\frac{x_i^{2n_i}}{n_{i\gamma}!(n_i!)^2}\left|\sum_{m_i=\max(0,n_{i\rho}-n_i)}^{\min(n_{i\rho},n_{i\gamma})}\begin{pmatrix} n_{i\gamma}\\ m_i\end{pmatrix}\begin{pmatrix} n_i\\ n_{i\rho}-m_i\end{pmatrix}\right.\right.\\
&\left.\left.\times R_i^{m_i}T_i^{n_{i\gamma}-m_i}t_i^{n_{i\rho}-m_i}r_i^{n_i-n_{i\rho}+m_i}\right|^2\right]\\
&\times\left|\langle\sqrt{\frac{(\sum_i(2n_i+n_{i\gamma}))!}{\prod_i[n_i!n_{i\rho}!(n_i+n_{i\gamma}-n_{i\rho})!]}}\prod_i\left({}^{n_i}_{\otimes}{}_i\tau^{\otimes\, n_{i\rho}}{}_i\rho^{\otimes\, n_i+n_{i\gamma}-n_{i\rho}}{}_i\sigma\right)_{sym},\right.\\
&\qquad\left.\prod_i\left({}^{n_i}_{\otimes}{}_i\tau^{\otimes\, n_{i\rho}}{}_i\rho^{\otimes\, n_i+n_{i\gamma}-n_{i\rho}}{}_i\sigma\right)_{sym}\rangle\right|^2.
\end{aligned}
\tag{A.12}
$$

The inner product in the last expression equals to [6]

$$
\sqrt{\frac{\prod_i[n_i!n_{i\rho}!(n_i+n_{i\gamma}-n_{i\rho})!]}{(\sum_i(2n_i+n_{i\gamma}))!}}.
$$

[4] The number of the 'particle' ${}_i\sigma$ in (A.10) is $n_i + n_{i\gamma} - n_{i\rho}$, setting the number of the 'particle' ${}_i\rho$ to $n_{i\rho}$.

[5] The range is obtained by inequalities $n_i \geq 0$, $0 \leq m_i \leq n_{i\gamma}$, $0 \leq l_i \leq n_i$, $l_i + m_i = n_{i\rho}$ and $n_i + n_{i\gamma} - n_{i\rho} \geq 0$.

[6] (A.11) is normalized to have unit norm.

Finally, by using (A.9) and exchanging the order of the summation suitably, we can obtain

$$
\begin{aligned}
&P(\{n_{i\rho}\}|\{n_{i\gamma}\}) \\
&= \prod_i \Big[ (1-x_i) x_i^{2n_{i\rho}} \left(1-|R_i|^2\right)^{n_{i\gamma}+n_{i\rho}} \\
&\times \sum_{l_i=0}^{\min(n_{i\gamma},n_{i\rho})} \sum_{m_i=0}^{\min(n_{i\gamma},n_{i\rho})} \frac{\left[-|R_i|^2/(1-|R_i|^2)\right]^{l_i+m_i} n_{i\gamma}! n_{i\rho}!}{l_i!(n_{i\gamma}-l_i)!(n_{i\rho}-l_i)! m_i!(n_{i\gamma}-m_i)!(n_{i\rho}-m_i)!} \\
&\times \sum_{n_i=n_{i\rho}-\min(l_i,m_i)}^{\infty} \frac{n_i!(n_i-n_{i\rho}+n_{i\gamma})!}{(n_i-n_{i\rho}+l_i)!(n_i-n_{i\rho}+m_i)!} (x_i^2|R_i|^2)^{n_i-n_{i\rho}} \Big] .
\end{aligned}
$$

This is what we had to show.

## A.2 A proof of Lemma 2

In this appendix we give a proof of Lemma 2.

**Proof**

Since a set of all ${}_i\tau$ and ${}_i\sigma$ generates $\mathcal{H}_{H^+}$ [67], the definition of $T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}}$ leads

$$
T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}} = \sum_{\{n_{i\sigma}\},\{n_{i\tau}\}} \langle \{n_{i\tau}, n_{i\rho}, n_{i\sigma}\}|S|\{n_{i\gamma}\}\rangle \langle \{n'_{i\gamma}\}|S|\{n_{i\tau}, n'_{i\rho}, n_{i\sigma}\}\rangle, \quad \text{(A.13)}
$$

where

$$
|\{n_{i\tau}, n_{i\rho}, n_{i\sigma}\}\rangle \equiv \prod_i \left[ \frac{1}{\sqrt{n_{i\tau}! n_{i\rho}! n_{i\sigma}!}} \left(a^\dagger({}_i\tau)\right)^{n_{i\tau}} \left(a^\dagger({}_i\rho)\right)^{n_{i\rho}} \left(a^\dagger({}_i\sigma)\right)^{n_{i\sigma}} \right] |0\rangle.
$$

In the expression, $S|\{n_{i\gamma}\}\rangle$ is given by (A.10) and $S|\{n'_{i\gamma}\}\rangle$ is obtained by replacing $n_{i\gamma}$ with $n'_{i\gamma}$ in (A.10). Now, those orthonormal basis vectors of the form $|\{n_{i\tau}, n_{i\rho}, n_{i\sigma}\}\rangle$ that have a non-zero inner product with $S|\{n_{i\gamma}\}\rangle$ must also be of the form (A.11). Thus, $T^{\{n_{i\gamma}\}\{n'_{i\gamma}\}}_{\{n_{i\rho}\}\{n'_{i\rho}\}}$ vanishes unless there exist such a set of integers $\{n_i, n'_i\}$ $(i = 1, 2, \cdots)$ that

$$
\begin{aligned}
n_i &= n'_i \\
n_i + n_{i\gamma} - n_{i\rho} &= n'_i + n'_{i\gamma} - n'_{i\rho} \quad \text{(A.14)}
\end{aligned}
$$

for $\forall i$. The existence of $\{n_i\}$ and $\{n'_i\}$ is equivalent to the condition $n_{i\gamma} - n'_{i\gamma} = n_{i\rho} - n'_{i\rho}$ for $\forall i$.

□

## A.3 On-shell brick wall model

When we performed the differentiation with respect to $\beta_\infty$ to obtain the total energy and the entropy in section 3.2, the surface gravity $\kappa_0$ of the black hole and the inverse temperature $\beta_\infty$ of gas on the black hole background were considered as independent quantities. Since in equilibrium these quantities are related by $\beta_\infty^{-1} = \kappa_0/2\pi$, we have imposed this relation, which we call the on-shell condition, after the differentiation. In fact, we have shown that the wall contribution to gravitational

energy is zero and the backreaction can be neglected, if and only if the on-shell condition is satisfied.

On the other hand, in the so-called on-shell method [82, 32, 75], the on-shell condition is implemented before the differentiation. Now let us investigate what we might call an on-shell brick wall model. With the on-shell condition, the wall contribution to the free energy of the scalar field considered in subsection 3.2.3 is calculated as

$$F_{wall}^{(on-shell)} = -\frac{A}{4}\frac{\beta_\infty^{-1}}{360\pi}\frac{1}{\alpha^2}. \tag{A.15}$$

If we define total energy and entropy in the on-shell method by

$$\begin{aligned} U_{wall}^{(on-shell)} &\equiv \frac{\partial}{\partial\beta_\infty}\left(\beta_\infty F_{wall}^{(on-shell)}\right), \\ S_{wall}^{(on-shell)} &\equiv \beta_\infty^2\frac{\partial}{\partial\beta_\infty}F_{wall}^{(on-shell)}, \end{aligned} \tag{A.16}$$

then these quantities can be calculated as

$$\begin{aligned} U_{wall}^{(on-shell)} &= 0, \\ S_{wall}^{(on-shell)} &= \frac{A}{4}\frac{1}{360\pi}\frac{1}{\alpha^2} = \frac{1}{4}S_{wall}, \end{aligned} \tag{A.17}$$

where $S_{wall}$ is the wall contribution (3.62) to entropy of the scalar field with $T_\infty = T_{BH}$.

It is notable that the total energy $U_{wall}^{(on-shell)}$ in the on-shell method is zero irrespective of the value of the cutoff $\alpha$. However, $S_{wall}^{(on-shell)}$ is always smaller than $S_{wall}$. It is because some physical degrees of freedom are frozen by imposing the on-shell condition before the differentiation. Thus, we might miss the physical degrees of freedom in the on-shell method.

## A.4 Symmetric property of the entanglement entropy for a pure state

In this appendix we first give an abstract expression for the reduced density operators $\rho_1$ and $\rho_2$ corresponding to a pure state $u$ in $\mathcal{F} = \mathcal{F}_1\bar{\otimes}\mathcal{F}_2$, which do not use the subtrace. Then with the help of them we prove that $S_{ent}$ obtained from $\rho_1$ and $\rho_2$ coincide with each other. We follow the notations in §3.3.1.

**Proposition 1** *For an arbitrary element $u$ of $\mathcal{F} = \mathcal{F}_1\bar{\otimes}\mathcal{F}_2$, there are antilinear bounded operators $A_u \in \bar{\mathcal{B}}(\mathcal{F}_1, \mathcal{F}_2)$ and $A_u^* \in \bar{\mathcal{B}}(\mathcal{F}_2, \mathcal{F}_1)$ such that*

$$(A_u x, y) = (A_u^* y, x) = (u, x \otimes y) \tag{A.18}$$

*for $^\forall x \in \mathcal{F}_1$ and $^\forall y \in \mathcal{F}_2$.*

*Proof.* Fix an arbitrary element $x$ of $\mathcal{F}_1$. Then $(u, x \otimes y)$ gives a linear bounded functional of $y(\in \mathcal{F}_2)$ since

$$|(u, x \otimes y)| \leq ||u||||x||||y||.$$

Hence by Riesz's theorem there is a unique element $z_{u,x}$ of $\mathcal{F}_2$ such that

$$(z_{u,x}, y) = (u, x \otimes y) \tag{A.19}$$

for $^{\forall}y \in \mathcal{F}_2$. Let us define $A_u$ by $A_u : x \to z_{u,x}$. It is evident that $A_u$ is an antilinear bounded operator from $\mathcal{F}_1$ to $\mathcal{F}_2$ since

$$||A_u x|| = ||z_{u,x}|| = ||u||||x||.$$

Exchanging the roles played by $\mathcal{F}_1$ and $\mathcal{F}_2$ in the above argument, it is shown that there is an antilinear bounded operator $A_u^*$ from $\mathcal{F}_2$ to $\mathcal{F}_1$ such that $(A_u^* y, x) = (u, x \otimes y)$. □

Note that $A_u$ and $A_u^*$ defined above are written as

$$\begin{aligned} A_u x &= \sum_j f_j (x \otimes f_j, u) \quad , \\ A_u^* y &= \sum_i e_i (e_i \otimes y, u) \quad . \end{aligned} \tag{A.20}$$

Using this expression, it is easily shown that

$$\begin{aligned} A_u^* A_u x &= \sum_{ij} e_i \left(e_i \otimes f_j, (u, x \otimes f_j) u\right) \quad , \\ A_u A_u^* y &= \sum_{ij} f_j \left(e_i \otimes f_j, (u, e_i \otimes y) u\right) \quad . \end{aligned} \tag{A.21}$$

These coincide with $\rho_1$ and $\rho_2$, respectively, if $u$ has unit norm (see Eq. (3.110) and (3.111)). Therefore the following proposition says that $\rho_1$ and $\rho_2$ have the same spectrum and the same multiplicity and that entropy of them are identical.

**Proposition 2** *$\rho_u^{(1)}$ $(\in \mathcal{B}(\mathcal{F}_1))$ and $\rho_u^{(2)}$ $(\in \mathcal{B}(\mathcal{F}_2))$ defined by*

$$\begin{aligned} \rho_u^{(1)} &= A_u^* A_u \quad , \\ \rho_u^{(2)} &= A_u A_u^* \end{aligned} \tag{A.22}$$

*are non-negative, trace-class self-adjoint operators, where $A_u$ and $A_u^*$ are defined in Proposition 1 for an arbitrary element $u$ of $\mathcal{F}$. The spectrum and the multiplicity of $\rho_u^{(1)}$ and $\rho_u^{(2)}$ are identical for all non-zero eigenvalues.*

*Proof.* In general

$$(x', \rho_u^{(1)} x) = (A_u x, A_u x')$$

for $^{\forall}x, x'$ $(\in \mathcal{F}_1)$ by definition. Therefore

$$(x, \rho_u^{(1)} x) = ||A_u x||^2 \geq 0 \tag{A.23}$$

and

$$\begin{aligned} \mathrm{Tr}_{\mathcal{F}_1}(\rho_u^{(1)}) &= \sum_i (A_u e_i, A_u e_i) \\ &= \sum_{i,j} (A_u e_i, f_j)(f_j, A_u e_i) \\ &= \sum_{i,j} |(u, e_i \otimes f_j)|^2 \\ &= ||u||^2, \end{aligned} \tag{A.24}$$

i.e. $\rho_u^{(1)}$ is non-negative and trace-class. In general a non-negative operator is self-adjoint and a trace-class operator is compact. Thus the eigenvalue expansion theorem for a self-adjoint compact operator says that all eigenvalues of $\rho_u^{(1)}$ are

discrete except zero and have finite multiplicity. For a later convenience let us denote the non-zero eigenvalues and the corresponding eigenspaces as $\lambda_i$ and $\mathcal{F}_{1,i}$ $(i = 1, 2, \cdots)$.

Similarly, it is shown that $\rho_u^{(2)}$ is non-negative and trace-class and that all eigenvalues of it are discrete except zero and have finite multiplicity.

Now $\ker \rho_u^{(2)} = \ker A_u^*$ since $\rho_u^{(2)} y = A_u A_u^* y$ and $(y, \rho_u^{(2)} y) = ||A_u^* y||^2$ for an arbitrary element $y$ of $\mathcal{F}_2$ by definitions. Moreover, from (A.18) it is evident that

$$y \perp \mathrm{Ran} A_u \Leftrightarrow y \in \ker A_u^*.$$

With the help of these two facts $\mathcal{F}_2$ is decomposed as

$$\mathcal{F}_2 = \ker \rho_u^{(2)} \oplus \overline{\mathrm{Ran} A_u}. \tag{A.25}$$

where the overline means to take a closure.

Moreover, it is easily shown by definitions that

$$\begin{aligned} \rho_u^{(2)} A_u x &= \lambda_i A_u x \quad , \\ (A_u x, A_u x') &= \lambda_i (x', x) \end{aligned} \tag{A.26}$$

for ${}^\forall x$ $(\in \mathcal{F}_{1,i})$ and ${}^\forall x'$ $(\in \mathcal{F}_{1,i'})$. Hence $A_u$ maps the eigenspace $\mathcal{F}_{1,i}$ to a eigenspace of $\rho_u^{(1)}$ with the same eigenvalue, preserving its dimension. Taking account of Eq.(A.25), this implies that the spectrum and the multiplicity of $\rho_u^{(1)}$ and $\rho_u^{(2)}$ are identical for all non-zero eigenvalues. □

## A.5 Entanglement energy for the case of $B = \mathrm{R}^2$ in Minkowski spacetime

First we take $B = \mathrm{R}^2$. Without loss of generality the resulting two half-spaces are represented as $\{(x_1, x_2, x_3) : x_1 > 0\}$ and $\{(x_1, x_2, x_3) : x_1 < 0\}$ [37].

Here some comments are in order. Since all the degrees of freedom on and across $B$, which is infinite, contribute to the entanglement energy, a suitable cut-off length $a(> 0)$ should be introduced to avoid the ultra-violet divergence. For the same reason, the infra-red divergence is also anticipated in advance, since $B$ is non-compact in this model. The latter is taken care of by considering the massive case since the inverse of the mass characterizes a typical size of the spreading of the field. Clearly $a$ should be taken short enough in the unit of the Compton length of the field, $m_\phi^{-1}$, to obtain meaningful results. Therefore we shall only pay attention to the leading order in the limit $m_\phi a \to 0$ in the course of calculation as well as in the final results. These remarks are valid in any model of this type, and the same remarks apply to the case of the entanglement entropy, too [37, 38].

In order to calculate $\langle : H_{tot} : \rangle_{\rho'}$ for the present case, we first note that the term $V_{AB} q^A q^B$ in Eq.(3.114) corresponds to the expression $\int \left[ (\vec{\nabla}\phi)^2 + m_\phi^2 \phi^2 \right] d^3x$ read off from Eq.(3.156), which defines an operator $V(x, y)$ acting on a space $\mathcal{W} = \left( \{\phi(\cdot)\}, d^3x \right)$. In order to use the formula (3.149), thus, we need the inverse of the positive square-root of $aV$. For this purpose, it is convenient to work in the momentum representation of $\mathcal{W}$ [37] given by

$$\begin{aligned} \phi(\vec{x}) &= \int \frac{d^3k}{(2\pi)^3} \, \phi_{\vec{k}} \exp[i\vec{k} \cdot \vec{x}], \\ \phi_{\vec{k}} &= \int d^3x \, \phi(\vec{x}) \exp[-i\vec{k} \cdot \vec{x}]. \end{aligned} \tag{A.27}$$

The results are

$$\begin{aligned} V(\vec{x},\vec{y}) &= \int \frac{d^3k}{(2\pi)^3}(\vec{k}^2+m_\phi^2)\exp[i\vec{k}\cdot(\vec{x}-\vec{y})] \quad , \\ W^{-1}(\vec{x},\vec{y}) &= \int_{\mathrm{R}^3} \frac{d^3k}{(2\pi)^3}(\vec{k}^2+m_\phi^2)^{-1/2}\exp[i\vec{k}\cdot(\vec{x}-\vec{y})]. \end{aligned} \tag{A.28}$$

Note that both $V(\vec{x},\vec{y})$ and $W^{-1}(\vec{x},\vec{y})$ are symmetric under the exchange of $\vec{x}$ and $\vec{y}$. (The cut-off must preserve this property.) Now the formula (3.149) gives

$$\begin{aligned} \langle : H_{tot} : \rangle_{\rho'} &= -\frac{1}{2}\int_{y_1<-a} d^3y \int_{x_1>a} d^3x \int_{|k_1|<a^{-1}} \frac{d^3k}{(2\pi)^3}(\vec{k}^2+m_\phi^2)\exp[i\vec{k}\cdot(\vec{y}-\vec{x})] \\ &\quad \times \int_{|k'_1|<a^{-1}} \frac{d^3k'}{(2\pi)^3}(\vec{k'}^2+m_\phi^2)^{-1/2}\exp[i\vec{k'}\cdot(\vec{x}-\vec{y})], \end{aligned} \tag{A.29}$$

where, as discussed above, a cut-off length $a$ was introduced in the integral.

Since the integrand is invariant under the translation along $B$, the integral with respect to $x_2$ and $x_3$ yields a divergent factor $A = \int_{R^2} dx_2 dx_3$. Clearly this factor should be interpreted as the area of $B$. If this divergent integral $A$ is factored out, we obtain the following convergent expression for $\langle : H_{tot} : \rangle_{\rho'}$:

$$\begin{aligned} \langle : H_{tot} : \rangle_{\rho'} &= -\frac{A}{2}\int_{-\infty}^{-a} dy_1 \int_a^\infty dx_1 \int \frac{d^2k_\parallel}{(2\pi)^2}\int_{-a^{-1}}^{a^{-1}} \frac{dk_1}{2\pi}\int_{-a^{-1}}^{a^{-1}} \frac{dk'_1}{2\pi} \\ &\quad \times(\vec{k_\parallel}^{\,2}+{k_1}^2+m_\phi^2)(\vec{k_\parallel}^{\,2}+{k'_1}^2+m_\phi^2)^{-1/2} \\ &\quad \times\exp[i(k_1-k'_1)(y_1-x_1)]. \end{aligned} \tag{A.30}$$

Here $\vec{k}_\parallel$ is a 2-vector lying along $B$ and $k_1$, $k'_1$ are components normal to $B$ (if we make an obvious identification of $\mathrm{R}^3$ with its Fourier space). Let us change the variables from $x_1$ and $y_1$ to $z \equiv x_1 - y_1$ and $u \equiv (x_1+y_1)/2$. Then $z$ and $u$ take values in the range $z \le 2a$ and $-(\frac{z}{2}-a) \ge u \ge (\frac{z}{2}-a)$, respectively. Hence the integration with respect to $u$ yields

$$\begin{aligned} \langle : H_{tot} : \rangle_{\rho'} &= -\frac{A}{2}\int_{2a}^\infty dz(z-2a)\int \frac{d^2k_\parallel}{(2\pi)^2}\int_{-a^{-1}}^{a^{-1}} \frac{dk_1}{2\pi}\int_{-\infty}^{\infty} \frac{dk'_1}{2\pi} \\ &\quad \times(\vec{k}_\parallel^2+{k_1}^2+m_\phi^2)(\vec{k}_\parallel^2+{k'_1}^2+m_\phi^2)^{-1/2}\exp[-i(k_1-k'_1)z] \\ &= -\frac{A}{2}\int_{2a}^\infty dz(z-2a)\int_{-a^{-1}}^{a^{-1}} \frac{dk_1}{2\pi}\cos(k_1z)\int_m^\infty \frac{d\kappa}{2\pi}\kappa(\kappa^2+{k_1}^2) \\ &\quad \times\int_{-\infty}^\infty \frac{dk'_1}{2\pi}(\kappa^2+{k'_1}^2)^{-1/2}\cos(k'_1z) \end{aligned} \tag{A.31}$$

in the leading order, where $\kappa$ is defined by $\kappa^2 = \vec{k}_\parallel^2 + m_\phi^2$. Here note that in this expression, the integration with respect to $k'_1$ followed by that with respect to $\kappa$ leads to an infra-red divergence if we set $m = 0$, in accordance with our discussion at the beginning of this subsection.

Now let us recollect some formulas with the modified Bessel functions [96]:

$$\begin{aligned} K_0(x) &= \int_0^\infty dt \frac{\cos t}{\sqrt{t^2+x^2}}, \\ \int_{x_0}^\infty dx\ xK_0(x) &= x_0K_1(x_0) \quad , \\ \int_{x_0}^\infty dx\ x^3K_0(x) &= x_0^3K_1(x_0)+2x_0^2K_2(x_0). \end{aligned} \tag{A.32}$$

With the help of these formulas $E^I_{ent}$ is written as

$$\begin{aligned} E^I_{ent} &= -\frac{A}{2}\int_{2a}^{\infty} dz(z-2a)\int_{-a^{-1}}^{a^{-1}} \frac{dk_1}{2\pi}\cos(k_1 z) \\ &\quad \times\frac{1}{2\pi^2}\left[\frac{m}{z}({k_1}^2+m_\phi^2)K_1(mz)+2\frac{m_\phi^2}{z^2}K_2(mz)\right], \\ &= \frac{A}{4\pi^3 a^3}[\alpha_1(ma)+\alpha_2(ma)+\alpha_3(ma)] \end{aligned} \tag{A.33}$$

in the leading order. Here we have introduced

$$\begin{aligned} \alpha_1(x) &\equiv -x\int_2^{\infty} d\xi\frac{\xi-2}{\xi^4}K_1(x\xi)\left[2\xi\cos\xi+(\xi^2-2)\sin\xi\right], \\ \alpha_2(x) &\equiv -2x^2\int_2^{\infty} d\xi\frac{\xi-2}{\xi^3}K_2(x\xi)\sin\xi, \\ \alpha_3(x) &\equiv -x^3\int_2^{\infty} d\xi\frac{\xi-2}{\xi^2}K_1(x\xi)\sin\xi. \end{aligned} \tag{A.34}$$

A numerical evaluation shows

$$[\alpha_1(x)+\alpha_2(x)+\alpha_3(x)]\sim 0.05 \quad \text{as} \quad x\to 0.$$

Therefore we get [7]

$$\langle : H_{tot} : \rangle_{\rho'} \approx \frac{0.05A}{4\pi^3 a^3}. \tag{A.35}$$

## A.6 States determined by variational principles

In this appendix we give derivations of formulas (3.203) and (3.209).

We consider a Hilbert space $\mathcal{F}$ of the form

$$\mathcal{F}=\mathcal{F}_1\otimes\mathcal{F}_2, \tag{A.36}$$

where $\mathcal{F}_1$ and $\mathcal{F}_2$ are Hilbert spaces with the same finite dimension $N$. An arbitrary unit element $|\phi\rangle$ of $\mathcal{F}$ is decomposed as

$$|\phi\rangle=\sum_{n=1}^{N}\sum_{m=1}^{N} C_{nm}|n\rangle_1\otimes|m\rangle_2, \tag{A.37}$$

where $|n\rangle_1$ and $|n\rangle_2$ $(n=1,2,\cdots,N)$ are orthonormal bases of $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and $\sum_{n,m}|C_{nm}|^2=1$ is understood. Here, without loss of generality, we can choose the orthonormal basis of $\mathcal{F}_2$ to be eigenstates of the normal-ordered Hamiltonian $:H_2:$ of the sub-system as

$$:H_2:|n\rangle_2 = E_n|n\rangle_2. \tag{A.38}$$

Since $C^\dagger C$ is a non-negative hermitian matrix, we can define a set of non-negative real numbers $\{p_n\}$, each of which is the eigenvalue of the matrix $C^\dagger C$. Hence,

$$C^\dagger C = V^\dagger P V, \tag{A.39}$$

[7] If we adopt another regularization scheme with the same cut-off length $a$, the result may change. However, the change is in sub-leading order.

where $V$ is a unitary matrix and $P$ is a diagonal matrix with diagonal elements $\{p_n\}$. With these definitions, the entanglement entropy $S_{ent}$ and the entanglement free energy $F_{ent}$ are calculated as

$$S_{ent} = -\sum_{n=1}^{N} p_n \ln p_n, \tag{A.40}$$

$$F_{ent} = \sum_{n=1}^{N}\sum_{m=1}^{N} E_n p_m |V_{mn}|^2 + T_{ent}\sum_{n=1}^{N} p_n \ln p_n. \tag{A.41}$$

The constraints $\sum_{n,m}|C_{nm}|^2 = 1$ and $V^\dagger V = \mathbf{1}$ are equivalent to

$$\begin{aligned} \sum_{n=1}^{N} p_n &= 1, \\ \sum_{l=1}^{N} V_{ln}^* V_{lm} &= \delta_{nm}. \end{aligned} \tag{A.42}$$

Thus, the variational principles are restated as follows: to maximize (A.40) under the constraints (A.42); to minimize (A.41) under the constraints (A.42).

Now, we shall show that expressions (A.40) and (A.41) are same as those for entropy and free energy in statistical mechanics in the subspace $\mathcal{F}_2$. Let us consider a density operator $\bar{\rho}$ on $\mathcal{F}_2$:

$$\bar{\rho} = \sum_{n=1}^{N}\sum_{m=1}^{N} \tilde{P}_{nm}|n\rangle_{22}\langle m|, \tag{A.43}$$

where $\tilde{P}_{nm}$ is an $N \times N$ non-negative hermitian matrix with unit trace. By diagonalizing the matrix $\tilde{P}$ as

$$\tilde{P} = \bar{V}^\dagger \bar{P} \bar{V}, \tag{A.44}$$

we obtain the following expressions for entropy and free energy.

$$\begin{aligned} S &= -\sum_{n=1}^{N} \bar{p}_n \ln \bar{p}_n, \\ F &= \sum_{n=1}^{N}\sum_{m=1}^{N} E_n \bar{p}_m |V_{mn}|^2 + T\sum_{n=1}^{N} \bar{p}_n \ln \bar{p}_n, \end{aligned} \tag{A.45}$$

where $\{\bar{p}_n\}$ are diagonal elements of the matrix $\bar{P}$ and $T$ is temperature. The constraints $\mathbf{Tr}\bar{\rho} = \mathbf{1}$ and $\bar{V}^\dagger \bar{V} = \mathbf{1}$ are restated as

$$\begin{aligned} \sum_{n=1}^{N} \bar{p}_n &= 1, \\ \sum_{l=1}^{N} \bar{V}_{ln}^* \bar{V}_{lm} &= \delta_{nm}. \end{aligned} \tag{A.46}$$

At this point, it is evident that the variational principles of maximum of entropy are the same in entanglement thermodynamics and statistical mechanics and that the principles of minimum of free energy are also the same in the two schemes. Hence, the principle of maximum of the entanglement entropy gives

$$\left(C^\dagger C\right)_{nm} = \frac{1}{N}\delta_{nm}, \tag{A.47}$$

as the principle of maximum of entropy gives the microcanonical ensemble

$$\tilde{P}_{nm} = \frac{1}{N}\delta_{nm} \tag{A.48}$$

in statistical mechanics. Similarly, the principle of minimum of the entanglement free energy gives

$$\left(C^\dagger C\right)_{nm} = Z^{-1}e^{-E_n/T_{ent}}\delta_{nm}, \tag{A.49}$$

as the principle of minimum of the free energy gives the canonical ensemble

$$\tilde{P}_{nm} = \bar{Z}^{-1}e^{-E_n/T}\delta_{nm} \tag{A.50}$$

in statistical mechanics, where $Z = \sum_n e^{-E_n/T_{ent}}$ and $\bar{Z} = \sum_n e^{-E_n/T}$ . It is easy to see that (A.47) and (A.49) are equivalent to (3.203) and (3.209), respectively, up to a unitary transformation in $\mathcal{F}_1$.

Finally we comment on the generalization of the analysis when the Hilbert space is divided into two subspaces with different dimensions ($dim\mathcal{F}_1 > dim\mathcal{F}_2$). In this case, by defining $S_{ent}$ from $\rho_2$, we obtain similar results.

## A.7 Bell states

In this appendix we show that in the finite dimensional case the orthonormal basis $\{|\psi_{nm}\rangle_1\}$ defined in the physical principle (c) is given by (3.213) uniquely up to a unitary transformation in $\mathcal{F}_{1+}$. After that, we derive the equation (3.214).

We consider the following decomposition of the Hilbert space $\mathcal{F}_1$.

$$\mathcal{F}_1 = \mathcal{F}_{1+} \otimes \mathcal{F}_{1-}, \tag{A.51}$$

where $\mathcal{F}_{1+}$ and $\mathcal{F}_{1-}$ are Hilbert spaces with the same finite dimension $N$. From the arguments in Appendix A.6, each of the basis $\{|\psi_{nm}\rangle_1\}$ is obtained by applying a unitary transformation in $\mathcal{F}_{1+}$ to the following state in $\mathcal{F}_1$.

$$|\phi\rangle_1 = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(|j\rangle_{1+} \otimes |j\rangle_{1-}\right), \tag{A.52}$$

where $|j\rangle_{1+}$ and $|j\rangle_{1-}$ $(j = 1, 2, \cdots, N)$ are orthonormal bases of $\mathcal{F}_{1+}$ and $\mathcal{F}_{1-}$, respectively. Evidently, any states given by (3.213) are obtained by this procedure. Moreover, it is easily confirmed as follows that a set of all states given by (3.213) is a complete orthonormal basis in the $N \times N$ dimensional Hilbert space $\mathcal{F}_1$.

$${}_1\langle\psi_{nm}|\psi_{n'm'}\rangle_1 = \frac{1}{N}\sum_{j=1}^{N} e^{2\pi ij(n'-n)/N}\delta_{mm'} = \delta_{nn'}\delta_{mm'}. \tag{A.53}$$

Let us suppose another complete orthonormal basis $\{|\bar{\psi}_{nm}\rangle_1\}$ in $\mathcal{F}_1$, each of which maximizes the entanglement entropy with respect to the decomposition $\mathcal{F}_1 = \mathcal{F}_{1+} \otimes \mathcal{F}_{1-}$. Since both $\{|\psi_{nm}\rangle_1\}$ and $\{|\bar{\psi}_{nm}\rangle_1\}$ are complete orthonormal basis in $\mathcal{F}_1$, they are related by a unitary transformation $U$ in $\mathcal{F}_1$. Moreover, $U$ is a unitary transformation in $\mathcal{F}_{1+}$, since any states maximizing the entanglement entropy are related by unitary transformations in $\mathcal{F}_{1+}$ as shown in Appendix A.6. Therefore, the orthonormal basis $\{|\psi_{nm}\rangle_1\}$ defined in the physical principle (c) is unique up to a unitary transformation in $\mathcal{F}_{1-}$ and is given by (3.213).

Now let us show the equation (3.214). The right hand side is transformed as follows.

$$
\begin{aligned}
&\frac{1}{N}\sum_{nm}|\psi_{nm}\rangle_1\otimes U_{nm}^{(2+)}|\tilde{\phi}_2\rangle_2 \\
&=\frac{1}{\sqrt{N}}\sum_{jk}\left(\frac{1}{N}\sum_{n}e^{2\pi i(j-k)n/N}\right)\times\sum_{mm'n'}{}_{2+}\langle k|n'\rangle_{2+}C_{n'm'} \\
&\times|(j+m)modN\rangle_{1+}\otimes|j\rangle_{1-}\otimes|(k+m)modN\rangle_{2+}\otimes|m'\rangle_{2-} \\
&=\sum_{mm'n'}\left(\frac{1}{\sqrt{N}}|(n'+m)modN\rangle_{1+}\otimes|(n'+m)modN\rangle_{2+}\right) \\
&\times C_{n'm'}|n'\rangle_{1-}\otimes|m'\rangle_{2-} \\
&=\left(\frac{1}{\sqrt{N}}\sum_{m''}|m''\rangle_{1+}\otimes|m''\rangle_{2+}\right) \\
&\times\left(\sum_{n'm'}C_{n'm'}|n'\rangle_{1-}\otimes|m'\rangle_{2-}\right). \qquad \text{(A.54)}
\end{aligned}
$$

The final expression is $|\phi\rangle$ itself.

# Bibliography


[1] M. Heusler, *Black Hole Uniqueness Theorems*, (Cambridge University Press 1996).

[2] R. Penrose, Riv. Nuovo Cimento **1**, 252 (1969); R. Penrose, in *General Relativity, an Einstein Centenary Survey*, ed. by S. W. Hawking and W. Israel (Cambridge University Press 1976).

[3] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, (Cambridge University Press 1973).

[4] J. D. Bekenstein, Lett. Nuovo Cimento **11**, 467 (1974); J. D. Bekenstein, in *Black Holes, Gravitational Radiation and the Universe*, edited by B. Bhawal and B. Iyer (Kluwer, Dordrecht 1998).

[5] K. Maeda, T. Tachizawa and T. Torii, Phys. Rev. Lett. **72**, 450 (1994).

[6] J. D. Bekenstein, Phys. Rev. D**7**, 2333 (1973).

[7] S. W. Hawking, Commun. math. Phys. **43**, 199 (1975).

[8] J. M. Bardeen, B. Carter and S. W. Hawking, Commun. Math. Phys. **31**, 161 (1973).

[9] P. Panangaden and R. M. Wald, Phys. Rev. **16**, 929 (1977).

[10] S. Mukohyama, Phys. Rev. **D56**, 2192 (1997).

[11] R. M. Wald, *General Relativity* (University of Chicago, Chicago, 1984).

[12] S. W. Hawking, Phys. Rev. Lett. **26**, 1344 (1971).

[13] I. Rácz and R. M. Wald, Class. Quantum Grav. **9**, 2643 (1992).

[14] W. Israel, Phys. Rev. Lett. **57**, 397 (1986).

[15] S. W. Hawking, Phys. Rev. **D14**, 2460 (1976).

[16] M. B. Green, J. H. Schwartz and E. Witten, *Superstring theory* (Cambridge University Press, 1987).

[17] J. Polchinski, *Superstring theory*, (Cambridge University Press 1998).

[18] J. Polchinski, S. Chaudhuri and C. V. Johnson, "Notes on D-Branes", hep-th/9602052 and references therein.

[19] A. Strominger and C. Vafa, Phys. Lett. **B379** (1996), 99.

[20] J. M. Maldacena, "Black Holes in String Theory", hep-th/9607235 and references therein.

[21] J. H. Schwarz, in *String Theory, Gauge Theory and Quantum Gravity*, Proceedings of the Spring School, Trieste, Italy, 1996, edited by R. Dijkgraaf *et al.* [Nucl. Phys. B (Proc. Suppl.) **55B**, 1 (1997)], and references therein.

[22] T. Banks, W. Fischler, S. H. Shenker and L. Susskind, Phys. Rev. **D55**, 5112 (1997).

[23] M. Li and E. Martinec, Class. Quantum Grav. **14**, 3187 (1997); **14**, 3205 (1997).

[24] T. Banks, W. Fischler, I. R. Klebanov and L. Susskind, Phy. Rev. Lett. **80**, 226 (1998); J. High Energy Phys. **01**, 008 (1998).

[25] C. Rovelli, "Loop Quantum Gravity", gr-qc/9710008 and references therein.

[26] A. Ashtekar and K. Krasnov, "Quantum Geometry and Black Holes", gr-qc/9804039 and references therein.

[27] V. P. Frolov and D. V. Fursaev, Class. Quant. Grav. **15**, 2041 (1998).

[28] G. W. Gibbons and S. W. Hawking, Phys. Rev. **D15**, 2752 (1977).

[29] S. W. Hawking and G. T. Horowitz, Phys. Rev. D**51**, 4302 (1995).

[30] J. D. Brown and J. W. York, Jr., Phys. Rev. D**47**, 1420 (1993).

[31] M. Banados, C. Teitelboim and J. Zanelli, Phys. Rev. Lett. **72**, 957 (1994).

[32] V. P. Frolov, D. V. Fursaev and A. I. Zelnikov, Phys. Rev. D**54**, 2711 (1996).

[33] R. M. Wald, Phys. Rev. **D48**, R3427 (1993).

[34] V. Iyer and R. M. Wald, Phys. Rev. **D52**, 4430 (1995).

[35] G. 'tHooft, Nucl. Phys. **B256**, 727 (1985).

[36] F. Pretorius, D. Vollick, and W. Israel, Phys. Rev. **D57**, (1998).

[37] L. Bombelli, R. K. Koul, J. Lee and R. D. Sorkin, Phys. Rev. **D34**, 373 (1986).

[38] M. Srednicki, Phys. Rev. Lett. **71**, 666 (1993).

[39] V. Frolov and I. Novikov, Phys. Rev. **D48**, 4545 (1993).

[40] S. Mukohyama, *On the Noether charge form of the first law of black hole mechanics*, gr-qc/9809050, to appear in Phys. Rev. **D**.

[41] S.A Hayward, S. Mukohyama and M. C. Ashworth, *Dynamical black-hole entropy*, gr-qc/9810006.

[42] S. Mukohyama and S. A. Hayward, in preparation.

[43] S. Mukohyama, Mod. Phys. Lett. **A11**, 3035 (1996).

[44] S. Mukohyama and W. Israel, Phys. Rev. **D**, 104005 (1998).

[45] S. Mukohyama, M. Seriu and H. Kodama, Phys. Rev. **D55**, 7666 (1997).

[46] S. Mukohyama, M. Seriu and H. Kodama, Phys. Rev. **D58**, 064001 (1998).

[47] S. Mukohyama, Phys. Rev. **D58**, 104023 (1998).

[48] V. Iyer and R. M. Wald, Phys. Rev. **D50**, 846 (1994).

[49] T. Jacobson, G. Kang and R. C. Myers, Phys. Rev. **D49**, 6587 (1994).

[50] R. M. Wald, J. Math. Phys. **31**, 2378 (1990).

[51] J. Lee and R. M. Wald, J. Math. Phys. **31**, 725 (1990).

[52] V. P. Frolov and D. N. Page, Phys. Rev. Lett. **71**, 3902 (1993).

[53] C. Truesdell, Rational Thermodynamics (Springer-Verlag 1984).

[54] H. Kodama, Prog. Theor. Phys. **63**, 1217 (1980).

[55] S. A. Hayward, Class. Quantum Grav. **15**, 3147 (1998).

[56] C. W. Misner and D. H. Sharp, Phys. Rev. **136**, B571 (1964).

[57] S. A. Hayward, Phys. Rev. **D53**, 1938 (1996).

[58] S. A. Hayward, Phys. Rev. **D49**, 6467 (1994).

[59] S. A. Hayward, *Dynamic wormholes*, gr-qc/9805019.

[60] R. Geroch, A. Held and R. Penrose, J. Math. Phys. **14**, 874 (1973); R. A. d'Inverno and J. Smallwood, Phys. Rev. **D22**, 1233 (1980); J. Smallwood, J. Math. Phys. **24**, 599 (1983); C. G. Torre, Class. Quantum Grav. **3**, 773 (1986); D. McMauns, Gen. Rel. Grav. **24**, 65 (1992); R. A. d'Inverno and J. A. G. Vickers, Class. Quantum Grav. **12**, 753 (1995).

[61] S. A. Hayward, Class. Quantum Grav. **10**, 779 (1993).

[62] S. W. Hawking, J. Math. Phys. **9**, 598 (1968).

[63] R. Penrose, Proc. R. Soc. London, **A381**, 53 (1982).

[64] W. H. Zurek and K. S. Thorne, Phys. Rev. Lett. **54**, 2171 (1985).

[65] R. M. Wald, "Black holes and thermodynamics", in *Black Hole Physics*, edited by V. De Sabbata and Z. Zhang (Kluwer, Boston 1992) pp 55-97.

[66] R. D. Sorkin, Phys. Rev. Lett. **56**, 1885(1986).

[67] R. M. Wald, Commun. math. Phys. **45**, 9(1975); Phys. Rev. **D13**, 3176(1976).

[68] G. Horowitz, J. Maldacena and A. Strominger, Phys. Lett. **B383**, 151 (1996).

[69] J. Maldacena and L. Susskind, Nucl. Phys. **B475**, 679 (1996).

[70] C. Callan and J. Maldacena, Nucl. Phys. **B475**, 645 (1996).

[71] M. Bershadsky, C. Vafa and V. Sadov, Nucl. Phys. **B463**, 398 (1996); A. Sen, Phys. Rev. **D53**, 2874 (1996).

[72] S. Das and S. Mathur, Nucl. Phys. **478**, 561 (1996).

[73] J. Maldacena and A. Strominger, Phys. Rev. **D55**, 861 (1997).

[74] F. Belgiorno and S. Liberati, Phys. Rev. **D53**, 3172 (1996); S. Liberati, Nuov. Cim. **112B**, 405 (1997).

[75] F. Belgiorno and M. Martellini, Phys. Rev. **D53**, 7073 (1996).

[76] L. Susskind and J. Uglum, Phys. Rev. **D50**, 2700 (1994).

[77] J. B. Hartle and S. W. Hawking, Phys. Rev. **D13**, 2188 (1976).

[78] D. G. Boulware, Phys. Rev. **D11**, 1404 (1975); C. J. Isham, "Quantum field theory in curved spacetime: an overview", in *Eighth Texas Symposium on Relativistic Astrophysics*, edited by M. D. Papagiannis (Ann. NY Acad. Sci., 1977) p. 114; P. C. W. Davies, "Thermodynamics of black holes", Rep. Prog. Phys. **41**, 1313 (1978); W. Israel, "Gedanken-experiments in black hole thermodynamics", in *Black Holes: Theory and Observation*, edited by F. Hehl, C. Kiefer and R. Metzler (Springer, Berlin 1998).

[79] R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon Press, Oxford 1934) p. 318.

[80] B. P. Jensen, J. G. McLaughlin and A. C. Ottewill, Phys. Rev. **D45**, 3002 (1992); P. R. Anderson, W. A. Hiscock and D. A. Samuel, Phys. Rev. **D51**, 4337(1995).

[81] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco 1973) p. 603.

[82] V. P. Frolov, Phys. Rev. Lett. **74**, 3319 (1995)

[83] C. Callan and F. Wilczek, Phys. Lett. **B333**, 55 (1994).

[84] J. L. F. Barbón, Phys. Lett. **B339**, 41 (1994); J. G. Demers, R. Lafrance and R. C. Myers, Phys. Rev. **D52**, 2245 (1995).

[85] W. Israel, Phys. Lett. **57A**, 106 (1976).

[86] H. Umezawa, Advanced Field Theory (AIP Press, New York 1993), Chap. 3.

[87] For example, G. Jumarie, *Relative information* (Springer, 1990).

[88] N. J. Cerf, and C. Adami, Phys. Rev. Lett. **79**, 5194 (1997).

[89] N. J. Cerf, and C. Adami, Phys. Rev. **A55**, 3371 (1997).

[90] C. H. Bennet, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[91] D. Bouwrester, J-W. Pan, K. Mattle, M Eibl, H. Weinfurter, and A. Zeilinger, Nature **390**, 575 (1997).

[92] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, Phys. Rev. Lett. **80**, 1121 (1998).

[93] L. Parker, Phys. Rev. **D12**, 1519 (1975).

[94] V. P. Frolov, D. V. Fursaev and A. I. Zelnikov, Nucl. Phys. **B486**, 229 (1996).

[95] S. W. Hawking, "Is Information Lost in Black Holes?", in *Black Holes and Relativistic Stars*, edited by R. M. Wald, (University of Chicago Press 1998).

[96] E.g., M. Abramowitz and I. A. Stegun (ed.), *Handbook of Mathematical Functions* (Dover, New York, 1972), Chapter 9.